

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**ANÁLISE DE SÉRIES TEMPORAIS FUZZY
PARA PREVISÃO E IDENTIFICAÇÃO DE
PADRÕES COMPORTAMENTAIS
DINÂMICOS**

FÁBIO JOSÉ JUSTO DOS SANTOS

ORIENTADORA: PROF. DRA. HELOISA DE ARRUDA CAMARGO

São Carlos - SP
Abril/2015

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**ANÁLISE DE SÉRIES TEMPORAIS FUZZY
PARA PREVISÃO E IDENTIFICAÇÃO DE
PADRÕES COMPORTAMENTAIS
DINÂMICOS**

FÁBIO JOSÉ JUSTO DOS SANTOS

Tese apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Doutor em Ciência da Computação, área de concentração: Inteligência Artificial
Orientadora: Dra. Heloisa de Arruda Camargo

São Carlos - SP
Abril/2015

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária/UFSCar**

S237as

Santos, Fábio José Justo dos.

Análise de séries temporais fuzzy para previsão e identificação de padrões comportamentais dinâmicos / Fábio José Justo dos Santos. -- São Carlos : UFSCar, 2015.
132 f.

Tese (Doutorado) -- Universidade Federal de São Carlos, 2015.

1. Inteligência artificial. 2. Séries temporais fuzzy. 3. Previsão. 4. Agrupamento dinâmico de séries temporais. 5. Pré-processamento. I. Título.

CDD: 006.3 (20^a)



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa de Pós-Graduação em Ciência da Computação

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Tese de Doutorado do candidato Fábio José Justo dos Santos, realizada em 30/04/2015:

Profa. Dra. Heloisa de Arruda Camargo
UFSCar

Prof. Dr. Orides Morandin Junior
UFSCar

Prof. Dr. Pablo Alberto Dalbem de Castro
IFSP

Profa. Dra. Roseli/Aparecida Francelin Romero
USP

Prof. Dr. Zhao Liang
FFCLRP/USP

Aos meus pais Fátima e Ladislau.

Ao meu amor, Priscila.

Ao meu grande amigo Benedito.

AGRADECIMENTOS

A Deus, por sempre estar junto de mim.

Aos meus pais, por terem me ensinado o valor da educação em todos os sentidos.

À minha esposa, pelo amor, incentivo, palavras de conforto e paciência.

Ao meu grande amigo Dito, que há muito tempo tem sido meu segundo pai.

À professora Heloisa, pela confiança, dedicação, paciência e orientação.

Aos membros da banca examinadora, pelos comentários, críticas e sugestões.

Aos meus irmãos de laboratório, que hoje fazem parte da minha vida: Dino, Pri, Cris, Mariana, Marcos, Edward e Helano.

Ao IFSP, por permitir que na reta final deste projeto eu pudesse me dedicar integralmente aos estudos.

A todos que de alguma forma contribuíram para que este trabalho fosse concluído.

"O sucesso nasce do querer, da determinação e persistência em se chegar a um objetivo. Mesmo não atingindo o alvo, quem busca e vence obstáculos, no mínimo fará coisas admiráveis"

José de Alencar

RESUMO

Os bons resultados obtidos pelas abordagens fuzzy utilizadas para a análise de séries temporais (ST) tem contribuído significativamente para o crescimento da área. Embora haja resultados satisfatórios na análise de ST com métodos que utilizam os conceitos clássicos de ST e também com os conceitos recentes de séries temporais fuzzy (STF), há uma carência de modelos que combinem ambas as áreas. Diante deste contexto, as contribuições deste trabalho estão associadas ao desenvolvimento de modelos para a análise de ST combinando os conceitos de STF e métodos estatísticos visando a melhora na acurácia das previsões e a identificação de alterações comportamentais nas séries. Com o objetivo de permitir uma melhor representação fuzzy dos valores *crisp* observados, as abordagens desenvolvidas nesta tese foram associadas a uma nova proposta de pré-processamento dos dados. A previsão de valores é calculada a partir de uma nova técnica de suavização combinada a uma extensão das relações lógicas fuzzy. Essa combinação permite que sejam considerados no cálculo do valor previsto diferentes graus de influência para o comportamento mais recente e para o comportamento mais antigo da série. Em ocasiões onde o modelo não dispõe do conhecimento necessário para o cálculo do valor previsto, os conceitos de regressão linear simples são associados aos conceitos das STF para identificar a tendência mais recente da ST. A abordagem desenvolvida para a análise comportamental das séries tem como objetivo identificar mudanças no comportamento a partir da definição de protótipos que representam um grupo de ST e da segmentação das séries a serem analisadas. Nesta nova abordagem, a dissimilaridade entre um segmento de uma ST e o intervalo correspondente de um determinado protótipo é definida por meio da métrica *Dynamic Time Warping* (DTW) Fuzzy, ponderada por uma nova técnica de suavização aplicada à matriz de distâncias entre os dados observados. A acurácia obtida pelo modelo de previsão não só comprova a eficácia da abordagem desenvolvida, como também demonstra a evolução do modelo ao longo da pesquisa e a importância do pré-processamento nas previsões. A análise das ST segmentadas identifica satisfatoriamente as alterações comportamentais das séries por meio do cálculo da pertinência dos segmentos nos respectivos grupos representados pelos protótipos.

Palavras-chave: Séries Temporais Fuzzy; Previsão; Agrupamento Dinâmico de Séries Temporais; Pré-processamento

ABSTRACT

The good results obtained by the fuzzy approaches applied in the analysis of time series (TS) has contributed significantly to the growth of the area. Although there are satisfactory results in TS analysis with methods that use the classic concepts of TS and with the recent concepts of fuzzy time series (FTS), there is a lack of models combining both areas. Face of this context, the contributions of this thesis are associated with the development of models for TS analysis combining the concepts of FTS with statistical methods aiming at the improvement in accuracy of forecasts and in identification of behavioral changes in the TS. In order to allow a suitable fuzzy representation of crisp values observed, the approaches developed in this thesis were combined with a new proposal for pre-processing of the data. The prediction value is calculated from a new smoothing technique combined with an extension of the fuzzy logic relationships. This combination allow to be considered in value computed different degrees of influence to the most recent behavior and to the oldest behavior of the series. In situations where the model does not have the necessary knowledge to calculate the predicted value, the concepts of simple linear regression are combined with the concepts of the FTS to identify the most recent trend in the TS. The approach developed for the behavioral analysis of the TS aims to identify changes in behavior from the definition of prototypes that represent the groups of the TS and from the segmentation of the series that will be analyzed. In this new approach, the dissimilarity between a segment of a TS and the corresponding interval of a given prototype is defined by metric Fuzzy Dynamic Time Warping weighted by a new smoothing technique applied to the distance matrix between the observed data. The accuracy obtained by the forecast model not only demonstrates the effectiveness of the developed approach, but also shows the evolution of model throughout the research and the importance of preprocessing in the forecast. The analysis of segmented TS identifies satisfactorily the behavioral changes of the series by calculating the membership functions of these segments in the respective groups represented by the prototypes.

Keywords: Fuzzy Time Series; Forecast; Dynamic Clustering of Time Series; Pre-processing.

LISTA DE FIGURAS

Figura 2.1 - Série temporal com tendência de crescimento	22
Figura 2.2 – Série temporal com comportamento sazonal	22
Figura 2.3 – Série temporal com comportamento irregular	23
Figura 2.4 – Gráfico de dispersão com reta regressora e parâmetros β_0 e β_1	25
Figura 2.5 – Retas de regressão para a distribuição de dados em duas ST	27
Figura 2.6 – Fatores de ponderação em uma ST com dez amostras para $\alpha = 0, 2$, $\alpha = 0, 5$ e $\alpha = 0, 8$	28
Figura 2.7 – Série temporal fuzzificada	32
Figura 2.8 – RLF extraídas da série temporal fuzzificada	33
Figura 2.9 – RLF de 3ª – <i>ordem</i>	33
Figura 3.1 - Possíveis cenários para o agrupamento de dados. (Adaptada de (THEODORIDIS; KOUTROUMBAS, 2009))	38
Figura 3.2 – Distância Euclidiana. Uma pequena variação de fase pode ocasionar uma distorção do resultado esperado	40
Figura 3.3 – Duas séries temporais analisadas pelo algoritmo DTW	43
Figura 3.4 – Matriz de distância entre os valores observados em duas ST com 5 amostras cada uma	43
Figura 3.5 – Preenchimento da matriz DTW	44
Figura 3.6 – Matriz DTW com o cálculo final da dissimilaridade entre duas ST	44
Figura 3.7 – Caminho DTW com distorção na análise	45
Figura 3.8 – Restrições de Sakoe-Chiba e de Itakura (Adaptada de (HENNIGER; MULLER, 2007))	46
Figura 3.9 – Diferentes centróides obtidos pelo <i>k-means</i>	49
Figura 3.10 – Rede dos vizinhos mais próximos para $k = 2$	53
Figura 3.11 – Mudança nos valores observados na série temporal.	54
Figura 3.12 – Mudança no formato da série temporal.	55
Figura 3.13 – Mudança nos valores e no formato da série temporal.	55
Figura 4.1 – Pré-processamento das ST	59
Figura 4.2 – Definição do domínio e dos conjuntos fuzzy conforme pré-processamento apresentado na Seção 4.1.2	62

Figura 4.3 – Exemplo de série temporal.....	64
Figura 4.4 – Comportamento dos graus de influência suavizados para $\theta < 1$ com 3 amostras.....	75
Figura 4.5 - Comportamento dos graus de influência suavizados para $\theta > 1$ com 3 amostras.....	76
Figura 4.6 – Reta de regressão para a previsão do índice TAIEX de 25 de novembro de 2002	80
Figura 5.1 – Etapas do processo de análise comportamental das ST	83
Figura 5.2 – Séries temporais da base ECG200 utilizadas para a definição dos protótipos (KEOGH, et al., 2011).....	85
Figura 5.3 – Protótipos das séries temporais definidos por meio do Fuzzy C-Means clássico utilizando a distância euclidiana	86
Figura 5.4 – Protótipos das séries temporais definidos por meio do Fuzzy C-Means utilizando a métrica DTW	86
Figura 5.5 – Evolução do processo de segmentação conforme abordagem apresentada	88
Figura 5.6 – Graus de influência das células da diagonal principal em uma matriz DTW para $p_0 = 0, 1$, $p_0 = 0, 4$ e $p_0 = 0, 8$	91
Figura 6.1 – Índices TAIEX reais e previstos para os meses de novembro e dezembro de 2002	98
Figura 6.2 – Índices TAEIX reais e previstos para os meses de novembro e dezembro de 2003	98
Figura 6.3 – Índices TAEIX reais e previstos para os meses de novembro e dezembro de 2004	98
Figura 6.4 – Índices TAIEX reais e previstos para os meses de novembro e dezembro de 2001	100
Figura 6.5 – Índices TAIEX reais e previstos para os meses de novembro e dezembro de 2002	101
Figura 6.6 – Índices TAIEX reais e previstos para os meses de novembro e dezembro de 2004	101
Figura 6.7 – Índices TAIEX reais e previstos para os meses de novembro e dezembro de 2001	103
Figura 6.8 – Índices TAIEX reais e previstos para os meses de novembro e dezembro de 2002	103
Figura 6.9 – Índices TAIEX reais e previstos para os meses de novembro e dezembro de 2003	104

Figura 6.10 – Índices TAIEX reais e previstos para os meses de novembro e dezembro de 2004	104
Figura 6.11 – Cenário onde ocorre um desvio pela previsão com o auxílio da regressão linear simples	106
Figura 6.12 – Quatro ST da base ECG200 segmentadas.....	109
Figura 6.13 – Comportamento da ST1 em relação aos protótipos do Grupo 1 e do Grupo 2	110
Figura 6.14 – Distância ponderada entre os conjuntos fuzzy que representam os dados observados entre as amostras 3 e 6 da ST1 e os protótipos	110
Figura 6.15 – Comportamento da ST2 em relação aos protótipos do Grupo 1 e do Grupo 2	112
Figura 6.16 – Comportamento da ST3 em relação aos protótipos do Grupo 1 e do Grupo 2	112
Figura 6.17 – Comportamento da ST4 em relação aos protótipos do Grupo 1 e do Grupo 2	113

LISTA DE TABELAS

Tabela 3.1 – Similaridade entre séries temporais	52
Tabela 4.1 – Valores <i>crisp</i> de representação para cada termo linguístico do índice TAIEX 2001	67
Tabela 4.2 – RLF de 2ª ordem pertencentes ao Grupo 1 da série temporal TAIEX 2001	67
Tabela 4.3 – RLF de 2ª ordem pertencentes ao Grupo 2 da série temporal TAIEX 2001	68
Tabela 4.4 – RLF de 2ª ordem pertencentes ao Grupo 3 da série temporal TAIEX 2001	68
Tabela 4.5 – Valores <i>crisp</i> de representação para os termos linguísticos do índice TAIEX 2003	72
Tabela 4.6 – Índices TAIEX 2003 fuzzificados para treinamento do modelo	73
Tabela 4.7 – Valores <i>crisp</i> de representação para os termos linguísticos do índice TAIEX 2002	79
Tabela 6.1 – Séries temporais com o número anual de matrículas	94
Tabela 6.2 – Número de matrículas anuais reais e previstas	95
Tabela 6.3 – Comparação da acurácia obtida pelas previsões com e sem o pré-processamento	95
Tabela 6.4 – Pré-processamento nas ST dos índices TAIEX entre 2001 e 2004	97
Tabela 6.5 – Índices TAIEX 2001 reais e previstos para os meses de novembro e dezembro	97
Tabela 6.6 – Índices TAIEX reais e previstos para os meses de novembro e dezembro de 2003	100
Tabela 6.7 – Pré-processamento nas ST após ajustes na definição dos termos linguísticos	102
Tabela 6.8 – Fatores de suavização utilizados nos experimentos	102
Tabela 6.9 – Raiz do erro médio quadrático das previsões para os índices TAIEX com diferentes abordagens	105
Tabela 6.10 – Segmentos em cada ST	108
Tabela 6.11 – Pertinência dos segmentos da ST1 nos protótipos do Grupo 1 e do Grupo 2	111
Tabela 6.12 – Comportamento das ST apresentadas em relação aos grupos	113

LISTA DE ABREVIATURAS E SIGLAS

- AES** – *Alisamento Exponencial Simples*
- ARIMA** - *Autoregressive Integrated Moving Average*
- CIF** – *Computational Intelligence in Forecasting*
- DTW** – *Dynamic Time Warping*
- EMQ** – *Erro Médio Quadrático*
- FCM** – *Fuzzy C-Means*
- IBERAMIA** – *Ibero-American Conference on Artificial Intelligence*
- IFSA** – *International Fuzzy Systems Association*
- MMQ** – *Método dos Mínimos Quadrados*
- RLF** – *Relações Lógicas Fuzzy*
- RMSE** – *Root Mean Square Error (Raiz do Erro Médio Quadrático)*
- SQR** – *Soma do Quadrado da Regressão*
- SQT** – *Soma do Quadrado Total*
- ST** – *Séries Temporais*
- STF** – *Séries Temporais Fuzzy*
- SVM** – *Support Vector Machine*
- STS** – *Short Time Series*
- TAIEX** – *Taiwan Stock Exchange Capitalization Weighted Stock Index*
- TME** – *Taxa Média de Erro*
- UCR** – *University of California, Riverside*

SUMÁRIO

CAPÍTULO 1 - INTRODUÇÃO	13
1.1 Motivação e objetivos	16
1.2 Organização do Trabalho	18
CAPÍTULO 2 - SÉRIES TEMPORAIS	20
2.1 Fundamentação teórica.....	21
2.2 Regressão linear simples	24
2.3 Alisamento exponencial simples	28
2.4 Segmentação de séries temporais	29
2.5 Séries Temporais Fuzzy.....	30
2.6 Pré-processamento em séries temporais fuzzy.....	34
2.7 Considerações finais	36
CAPÍTULO 3 - AGRUPAMENTO DE SÉRIES TEMPORAIS	37
3.1 Métricas para séries temporais	39
3.1.1 Distância Euclidiana	39
3.1.2 Distância de Hamming	40
3.1.3 Coeficiente de correlação de Pearson	40
3.1.4 <i>Short Time Series</i>	41
3.1.5 Similaridade triangular ou similaridade do cosseno	42
3.1.6 <i>Dynamic Time Warping</i>	42
3.2 Agrupamento de séries temporais.....	47
3.2.1 K-means.....	48
3.2.2 Fuzzy C-Means	50
3.2.3 Vizinho mais próximo	52
3.2.4 Classificação comportamental das séries temporais.....	54
3.3 Considerações finais	56
CAPÍTULO 4 - ANÁLISE DE SÉRIES TEMPORAIS FUZZY PARA PREVISÃO DE VALORES	57

4.1 Pré-processamento das séries temporais fuzzy	58
4.1.1 Pré-processamento: primeira abordagem	59
4.1.2 Pré-processamento: segunda abordagem	61
4.1.3 Pré-processamento: terceira abordagem	62
4.2 Modelos de previsão	64
4.2.1 Previsões em STF baseadas no intervalo temporal das amostras.....	65
4.2.2 Previsão em STF por meio de uma extensão do alisamento exponencial simples	69
4.2.3 Um modelo híbrido de previsão de valores combinando séries temporais fuzzy, regressão linear e uma nova técnica de suavização.....	74
4.3 Considerações finais	81
CAPÍTULO 5 - ANÁLISE DO COMPORTAMENTO DINÂMICO DAS STF.....	82
5.1 Definição dos protótipos	83
5.2 Segmentação das séries temporais	87
5.3 Agrupamento dinâmico das séries temporais.....	89
5.3.1 Algoritmo DTW fuzzy ponderado para cálculo da dissimilaridade	89
5.3.2 Cálculo da pertinência do segmento de uma ST nos grupos representados pelos protótipos	92
5.4 Considerações finais	92
CAPÍTULO 6 - EXPERIMENTOS E AVALIAÇÃO DAS ABORDAGENS DESENVOLVIDAS.....	93
6.1 Experimentos para validação do pré-processamento	93
6.2 Experimentos com previsão	96
6.2.1 Resultados e análise dos experimentos com previsões baseadas no intervalo temporal das amostras.....	96
6.2.2 Resultados e análise dos experimentos com previsões baseadas na combinação de STF e o alisamento exponencial simples	99
6.2.3 Resultados e análise dos experimentos com previsões realizadas a partir da combinação de uma nova técnica de alisamento, dos conceitos de STF e da regressão linear simples	101
6.3 Experimentos com agrupamento de séries temporais	107
6.4 Considerações finais	114

CAPÍTULO 7 - CONCLUSÕES E TRABALHOS FUTUROS.....	115
7.1 Contribuições	117
7.2 Publicações	118
7.3 Limitações e trabalhos futuros.....	119
REFERÊNCIAS.....	122
APÊNDICE A: TERMOS LINGÜÍSTICOS UTILIZADOS NA ANÁLISE	
COMPORTAMENTAL DAS ST ECG200 (KEOGH, et al., 2011)	131

Capítulo 1

INTRODUÇÃO

Considerada uma área relativamente nova ao ser comparada com pesquisas equivalentes da área de estatística, as séries temporais fuzzy (STF) foram introduzidas pela primeira vez por Song e Chissom (1993a, 1993b, 1994). Os conceitos de conjuntos fuzzy apresentados por Zadeh (1965) aplicados na representação linguística de séries temporais (ST), possibilitam o uso do raciocínio aproximado (PEDRYCZ; GOMIDE, 1998) durante o processo de inferência, facilitando assim o desenvolvimento de soluções mais adequadas a diversos tipos de problemas.

O primeiro experimento com STF para previsão de valores realizado por Song e Chissom (1993b), teve como objetivo prever o número anual de matrículas na Universidade do Alabama entre os anos de 1971 e 1992. No experimento, a partir da fuzzificação dos valores *crisp*¹ da ST, a base de conhecimento é definida por meio de relações lógicas fuzzy (RLF) que representam as relações de implicação dos conjuntos fuzzy ou termos linguísticos. Desde então, os bons resultados obtidos pelas STF em experimentos com poucas amostras disponíveis, dados incompletos, com

¹ Valor real observado

comportamento irregular, ou ainda, com dados representados por meio de termos linguísticos, tem contribuído significativamente para o crescente interesse de pesquisa pelas séries temporais fuzzy.

Um requisito importante para uma boa acurácia nas previsões em STF é o pré-processamento dos dados observados (SANTOS; CAMARGO, 2013). A definição adequada das funções de pertinência e do número total de conjuntos fuzzy que representarão os dados fuzzificados, possui influência direta na acurácia dos valores previstos.

Diversas abordagens têm contribuído para a evolução da área e, conseqüentemente, para a análise de problemas nos mais diversos cenários do mundo real. Em (MORAES, et al., 2013) um sistema de inferência multicamadas baseado em regras fuzzy é apresentado para a previsão da demanda em uma subestação de energia. Um modelo de previsão que combina *support vector machine* (SVM), redes neurais evolucionárias e regras fuzzy foi apresentado por Stepnicka, et al. (2013) e validado utilizando diversos conjuntos de dados de ST, dentre eles, o número de passageiros de uma companhia aérea internacional entre os anos de 1949 e 1960, o número de venda de carros na cidade de Quebec, Canadá, entre 1960 e 1968 e a produção mensal de leite entre os anos de 1962 e 1975. Dentre todas as áreas de aplicação de ST, certamente a área financeira é a mais explorada em relação aos experimentos para previsão de valores e, por esta razão, a que apresenta maior diversidade de modelos de previsão (CHEN; CHEN, 2014) (CHEN, et al., 2012) (CHEN; KAO, 2013) (SANTOS; CAMARGO, 2014a) (SANTOS; CAMARGO, NO PRELO) (USLU, et al., 2014) (WANG; LEE, 2014) e (ZHANG, 2003).

Considerando a área de estatística, as ST possuem diversas classificações de acordo com o enfoque desejado. A primeira delas é em relação a quantidade de variáveis observadas, podendo ser séries monovariáveis ou multivariáveis. Outra classificação possível é em relação ao comportamento dos valores. Quando os valores observados ficam em torno de uma média, estas ST são consideradas estacionárias. As séries não estacionárias apresentam tendências lineares positivas ou negativas. Uma ST ainda pode ser classificada como determinística, cujos valores podem ser obtidos por meio de uma função matemática sem a adição de um resíduo ou termo aleatório de ajuste, ou ainda não determinísticas, também conhecidas como estocásticas, onde

há a necessidade de considerar o termo aleatório para a representação dos valores observados ao longo da ST (BUENO, 2008).

Dentre os modelos clássicos para previsão de valores, merecem destaque o *Autoregressive Integrated Moving Average* (ARIMA) (BOX; JENKINS, 1976), que é um modelo auto-regressivo de média móvel, o Alisamento Exponencial Simples (BROWN; MEYER, 1961), que realiza o cálculo da previsão de valores por meio da suavização exponencial das amostras, sendo possível dar maior peso às amostras mais recentes ou mais antigas, e a Regressão Linear Simples (GALTON, 1894), modelo de previsão para séries temporais que considera a tendência linear da série para realizar a previsão de um valor.

Atualmente um dos principais desafios de previsão em séries temporais fuzzy é na combinação dos conceitos fuzzy com outras abordagens da área de inteligência computacional e da área de estatística para a obtenção de uma melhor acurácia nos valores previstos.

Outra importante linha de pesquisa está relacionada ao padrão comportamental das séries (D'URSO, et al., 2013) (HENSMAN, et al., 2015) e (THINH; ANH, 2012). Entende-se por padrão comportamental de uma ST, a identificação de tendências de crescimento ou diminuição nos valores reais, o comportamento sazonal apresentado ao longo do tempo, ou ainda, o comportamento irregular da série de modo a não permitir uma classificação. Assim como ocorre nos modelos convencionais, os métodos para agrupamento de ST tem como objetivo colocar no mesmo *cluster* séries temporais com padrões comportamentais semelhantes, e em *clusters* diferentes ST com comportamentos distintos. Usualmente duas são as abordagens utilizadas para o agrupamento das ST: baseada na extração de atributos e baseada na similaridade. Na primeira abordagem um conjunto de atributos é extraído da ST para identificá-la e, a partir destes atributos, o agrupamento é realizado. Na segunda abordagem, uma métrica para calcular a similaridade ou dissimilaridade entre séries temporais é utilizada em conjunto com um algoritmo de agrupamento (RATANAMAHATANA, et al., 2010).

A abordagem para análise de ST desenvolvida nesta tese é composta por três frentes de pesquisa. A primeira delas tem foco no pré-processamento das séries visando uma definição adequada dos termos linguísticos que representarão os valores

numéricos observados. A segunda frente de pesquisa, com foco na previsão de valores, visa a combinação de modelos estatísticos com os conceitos de STF associados a uma nova técnica de alisamento para a melhora da acurácia. A terceira frente busca identificar mudanças comportamentais em uma série temporal ao longo de um período de observação. Esta identificação é realizada por meio da definição de protótipos que representem os grupos de ST, por meio da segmentação das séries e, na última etapa, por meio do cálculo da dissimilaridade dos segmentos com os respectivos protótipos que representam o comportamento de um determinado grupo de séries temporais.

1.1 Motivação e objetivos

Ao analisar os artigos de séries temporais fuzzy publicados nos últimos anos, é possível identificar um crescimento expressivo em qualidade e em volume de publicações. Entretanto, um dos principais questionamentos que surgiram no início das pesquisas deste trabalho foi: por que investir no desenvolvimento de séries temporais fuzzy se a área de estatística apresenta métodos bastante consolidados para lidar com séries temporais?

Dentre os principais motivos que contribuíram para a decisão de desenvolver a pesquisa, merecem destaque a boa acurácia apresentada pelas STF, o baixo custo computacional de processamento quando comparado com os modelos estatísticos e a facilidade de modelar matematicamente os problemas temporais com o uso de termos linguísticos. Outro fator que contribuiu consideravelmente para que esta pesquisa fosse desenvolvida, é o quão promissora a abordagem fuzzy para tratamento de séries temporais tem se mostrado por meio das pesquisas com foco em RLF e seus resultados recentes, sinalizando que, embora tenhamos resultados significativos com os modelos disponíveis, ainda há muito a ser desenvolvido.

Uma característica comum dos trabalhos pesquisados são as abordagens independentes para o pré-processamento dos dados, para a previsão de valores e para o agrupamento das ST. Apesar do pré-processamento nos dados possuir influência direta na acurácia das previsões e na qualidade do agrupamento obtido, uma

quantidade significativa de trabalhos com STF pode ser encontrada na literatura sem a realização de nenhuma forma de pré-processamento.

Embora haja métodos satisfatórios para a previsão de valores tanto na área de estatística quanto na área de STF foi identificada uma carência de modelos de previsão que combinem o conhecimento das duas áreas. Assim, uma hipótese levantada nesta tese é que a combinação de conceitos já consolidados em ambas as áreas em uma só abordagem de previsão possa contribuir para uma melhor acurácia dos valores previstos em diferentes ST.

Em relação ao agrupamento de ST, a grande maioria dos métodos disponíveis consideram que uma série pertence ao mesmo *cluster* durante todo seu período de observação, o que pode ocasionar uma interpretação equivocada do comportamento da ST ao analisá-la. Desta forma a principal motivação para o desenvolvimento da abordagem para análise comportamental dinâmica é a possibilidade de uma série temporal associada a um grupo passar a apresentar um comportamento característico das ST pertencentes a outro grupo.

Diante deste contexto, este trabalho tem como objetivo tratar problemas de previsão e agrupamento de séries temporais, ambos combinados com o pré-processamento dos dados observados. Para isso foi desenvolvido um modelo de previsão que integra os conceitos estendidos de STF a uma nova proposta de alisamento e aos conceitos de regressão linear simples. O modelo desenvolvido para a análise comportamental tem como característica realizar uma análise dinâmica das séries onde, em diferentes intervalos de observação, uma mesma ST pode apresentar padrões comportamentais de diferentes grupos.

A partir da distribuição dos dados observados ao longo da ST, o pré-processamento tem como propósito identificar e remover os *outliers*, definir o universo de discurso, determinar a quantidade ideal de conjuntos fuzzy para representar adequadamente a ST e definir as funções de pertinência destes conjuntos.

Com o objetivo de melhorar a acurácia na predição de valores por meio das STF neste trabalho foram desenvolvidas três abordagens de previsão. A principal abordagem desenvolvida apresenta uma nova proposta de extensão das RLF para a representação do conhecimento onde são considerados 3 termos linguísticos na implicação das relações. Esta nova técnica para a representação do conhecimento é

combinada com um novo modelo de alisamento capaz de atribuir maior influência no cálculo do valor previsto ao comportamento mais recente, ou ao comportamento mais antigo da ST. Para auxiliar na previsão em situações onde a base de conhecimento não possui informações disponíveis para realizar o cálculo, o modelo de regressão linear simples é combinado com a representação fuzzy das amostras para a identificação da tendência mais recente dos dados observados.

A abordagem para agrupamento das ST apresentada nesta tese considera que as ST podem apresentar comportamento dinâmico e, por esta razão, não pertencerem ao mesmo *cluster* durante todo o período de observação dos valores. Para realizar uma análise onde seja possível identificar a qual *cluster* pertence um determinado período de observação, as ST foram segmentadas e cada segmento é avaliado de modo independente. Para definir a similaridade entre estes segmentos, foi introduzida uma extensão da métrica *Dynamic Time Warping* (DTW) (BELLMAN, 1957) com a restrição de Itakura (ITAKURA, 1975), onde os conjuntos fuzzy são considerados no cálculo da distância e os valores obtidos são ponderados por meio de uma nova proposta de alisamento. Por fim, a pertinência dos segmentos nos respectivos protótipos é definida com o auxílio do cálculo de pertinência do algoritmo Fuzzy C-Means (FCM).

1.2 Organização do Trabalho

O restante desta tese está organizado como segue:

No Capítulo 2 são apresentados os conceitos sobre ST e STF com foco no pré-processamento e nos modelos de previsão para ST clássicas e fuzzy, além dos conceitos relacionados aos métodos de segmentação das séries.

No Capítulo 3 são apresentadas as principais abordagens para agrupamento de ST e é realizada uma introdução as principais métricas de similaridade e dissimilaridade para análise das séries.

No Capítulo 4 é descrito o modelo de previsão desenvolvido, desde a abordagem do pré-processamento até o cálculo dos valores a serem previstos. Com o objetivo de fornecer subsídios para uma análise da evolução da pesquisa, no Capítulo 4

também são apresentados alguns dos métodos de previsão desenvolvidos ao longo do desenvolvimento da tese.

No Capítulo 5 é apresentada a abordagem para análise do comportamento dinâmico das séries temporais. Neste capítulo é apresentada uma nova técnica para o cálculo da dissimilaridade entre as ST que realiza a análise segmentada das mesmas.

No Capítulo 6 são apresentados os experimentos de previsão e de agrupamento das séries temporais, e uma análise é realizada com base nos resultados obtidos.

Por fim, no Capítulo 7 são apresentadas as conclusões, as contribuições e as limitações das abordagens desenvolvidas, além das possibilidades de expansão dos modelos em trabalhos futuros.

Capítulo 2

SÉRIES TEMPORAIS

Uma ST pode ser definida como uma coleção de dados numéricos obtidos com uma determinada frequência de tempo. Certamente as ST são o meio mais utilizado para a previsão de valores pois, implicitamente, já fornecem um meio de ponderação dos dados para o cálculo de valores futuros, além de permitirem a identificação de modelos estruturais com componentes sazonais ou tendências nos valores observados.

Como a maior parte dos procedimentos estatísticos foi desenvolvida para a análise de observações independentes, o estudo de séries temporais acaba por exigir o uso de técnicas específicas (EHLERS, 2009). A análise de ST pode ser feita no domínio temporal ou no domínio de frequência. No primeiro caso modelos paramétricos são construídos a partir de algum tipo de distribuição de probabilidade e geralmente utilizados para gerar previsões ou descrever o comportamento de uma série. A análise no domínio de frequência consiste em manipular a ST de modo a extrair o espectro da frequência. Suas principais aplicações estão no campo física e engenharia e, dentre outros objetivos, visam identificar uma periodicidade relevante na sequência amostral (MORETTIN; TOLOI, 1981). O foco deste trabalho está na análise no domínio temporal e os conceitos relacionados para a elaboração desta tese são apresentados neste capítulo. Maiores informações a respeito da análise no domínio de

frequência podem ser encontradas em (JENKINS; WATTS, 1968) (KOOPMANS, 1974) (MORETTIN, 1979).

Na Seção 2.1 são apresentados alguns conceitos básicos e definições a respeito de ST. O modelo de regressão linear simples utilizado como ferramenta de apoio na segmentação das ST e na previsão de valores é apresentado na Seção 2.2. A Seção 2.3 apresenta o alisamento exponencial simples, inspiração para o modelo de alisamento apresentado na Seção 4.2.3. A Seção 2.4 apresenta as principais técnicas de segmentação de ST. Os conceitos de STF, suas definições e como a extração do conhecimento é realizada, são descritos na Seção 2.5. Os conceitos base relacionados ao pré-processamento das STF são apresentados na 2.6. As considerações finais são apresentadas na Seção 2.7.

2.1 Fundamentação teórica

Os valores observados em uma ST podem representar dados discretos ou contínuos. Nas ST discretas as observações são realizadas em períodos de tempo, normalmente, equidistantes. A representação dos dados no tempo t em uma ST, definido o conjunto $T = \{t_1, t_2, \dots, t_n\}$, é denotada por $\{Y_t : t \in T\}$ (SOUZA, 1989). Para as ST contínuas, a partir da definição do conjunto $T = \{t : t_1 < t < t_2\}$, sua representação é dada por $\{Y(t) : t \in T\}$ (GRANGER; NEWBOLD, 1977).

De acordo com Pucciarelli (2005), uma série temporal é considerada estacionária se não há grandes variações no comportamento da média e da variância dos valores observados ao longo do tempo, refletindo de alguma forma um equilíbrio estável e um comportamento regular da sequência dos valores. Segundo Newbold (1994), caso a ST não tenha um padrão comportamental estacionário e o período observado for suficientemente longo, é possível identificar três comportamentos distintos associados: as tendências dos dados ao longo da série, a sazonalidade e a irregularidade.

No primeiro caso, os valores apresentam uma tendência a crescer ou decrescer ao longo do tempo. A ST pode apresentar um comportamento regular ou irregular na média e covariância, desde que mantida a tendência de crescimento ou atenuação

dos valores observados. A Figura 2.1 apresenta um exemplo de ST com tendência de crescimento nos valores.

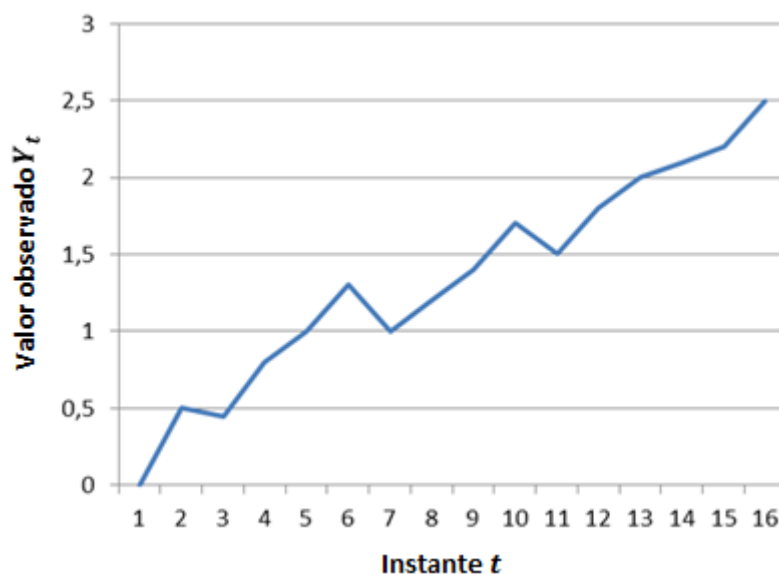


Figura 2.1 - Série temporal com tendência de crescimento

A sazonalidade em ST é caracterizada pelo comportamento da série possuir uma relação direta entre os dados observados e o tempo em si. Desta forma, o efeito sazonal é caracterizado pelos valores da ST assumirem comportamentos semelhantes em intervalos de tempo bem definidos. A Figura 2.2 apresenta um exemplo de ST com comportamento sazonal (NEWBOLD, 1994).

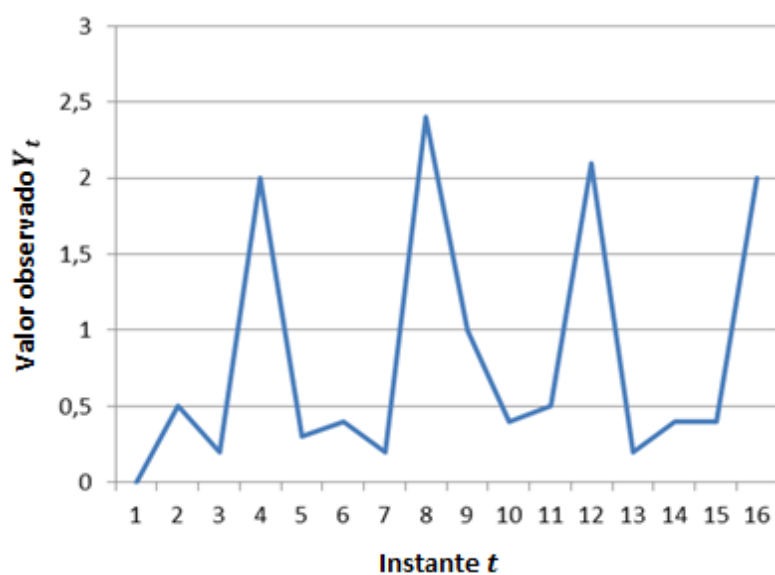


Figura 2.2 – Série temporal com comportamento sazonal

A sazonalidade de uma ST pode ser aditiva ou multiplicativa. No caso da sazonalidade aditiva, suponha que o aumento esperado na venda de chocolates no período da páscoa seja de 1 milhão de reais relação à média anual. Então, para levar em conta esta flutuação sazonal, as previsões para o período da páscoa dos próximos anos deve somar a quantia de 1 milhão de reais à média de vendas anual sem páscoa. No caso da sazonalidade multiplicativa, o aumento esperado é expresso em porcentagem, ou seja, considerando o exemplo da páscoa, ao afirmarmos que as vendas crescerão sessenta por cento, em números absolutos, o valor acrescido no período vai depender do comportamento da ST durante todo o período anterior, ou seja, se as vendas de chocolate forem pequenas durante o ano, o valor acrescido também será pequeno. De modo análogo, se os valores das vendas forem expressivos, podemos afirmar que a variação das vendas no período sazonal também será (EHLERS, 2009).

Nas ST com comportamento irregular, pouco ou nenhum padrão comportamental é identificado. A Figura 2.3 apresenta um exemplo de comportamento irregular de uma ST (NEWBOLD, 1994).

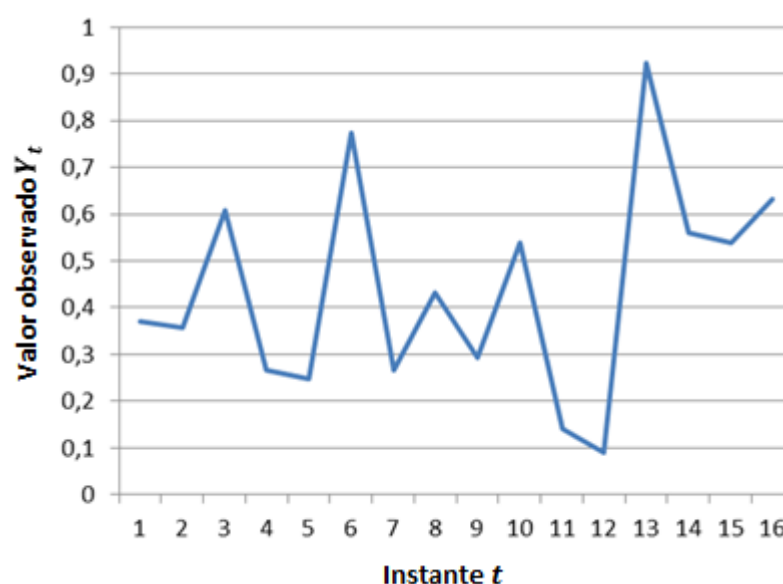


Figura 2.3 – Série temporal com comportamento irregular

Em muitas aplicações estatísticas, a ST representada graficamente é uma transformação dos dados originais que preserva as características dos dados originais (EHLERS, 2009). Dentre os métodos mais utilizados, podemos citar a transformação logaritma e a transformada Z e, em ambos os casos, os valores de previsão obtidos para a série transformada devem ser convertidos novamente para a escala original.

Há basicamente duas razões para a transformação dos dados observados: estabilizar a variância ao longo do período de observação da série, transformando uma ST não estacionária em estacionária, ou então tornar o efeito sazonal aditivo (MORETTIN; TOLOI, 1981). Entretanto, para a análise das ST com o objetivo de gerar previsões, Granger e Newbold (1976) e Makridakis e Hibon (1979) concluíram que a transformação dos dados originais não contribui para a melhora da acurácia nas previsões.

2.2 Regressão linear simples

A regressão linear simples consiste em uma análise estatística para verificar a existência de relação entre uma variável dependente Y , ou variável resposta, e uma variável independente X , ou variável regressora (PETERNELLI, 2004). Por meio desta análise é definido um modelo matemático que expressa, de modo linear, a relação existente entre as variáveis dependente e independente.

Proposto inicialmente por Galton (1894), o modelo matemático definido pela regressão linear simples tem como objetivo definir uma equação que represente a menor distância acumulada entre os pontos plotados no gráfico e a linha definida pela equação. Em outras palavras, a equação deve minimizar a soma dos desvios entre y_i e \hat{y}_i , onde y_i representa o valor observado e \hat{y}_i é o valor estimado pela equação de regressão linear simples. O modelo matemático completo, incluindo o valor residual, é apresentado na Equação 2.1.

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t, \quad t = 1, 2, \dots, n \quad (2.1)$$

onde y_t representa o valor projetado pela equação para a variável dependente no instante t , x_t contém o valor da variável independente, ε_t é a variável aleatória residual que representa o comportamento aleatório em Y que não pode ser explicado linearmente pelo comportamento de X . O parâmetro β_0 é o coeficiente linear da reta, também conhecido como intercepto, que contém o valor em Y para $x = 0$, ou seja, o ponto em que a reta regressora corta o eixo Y quando $x = 0$. O parâmetro β_1 do modelo é o coeficiente angular da reta, que representa a inclinação da reta regressora em relação ao eixo X , ou seja, a cada aumento de uma unidade em X , β_1 indica, por meio de um

valor real, a mudança média da distribuição de probabilidade em Y . A Figura 2.4 apresenta um gráfico de dispersão com sua respectiva reta regressora e a ilustração dos parâmetros β_0 e β_1 .

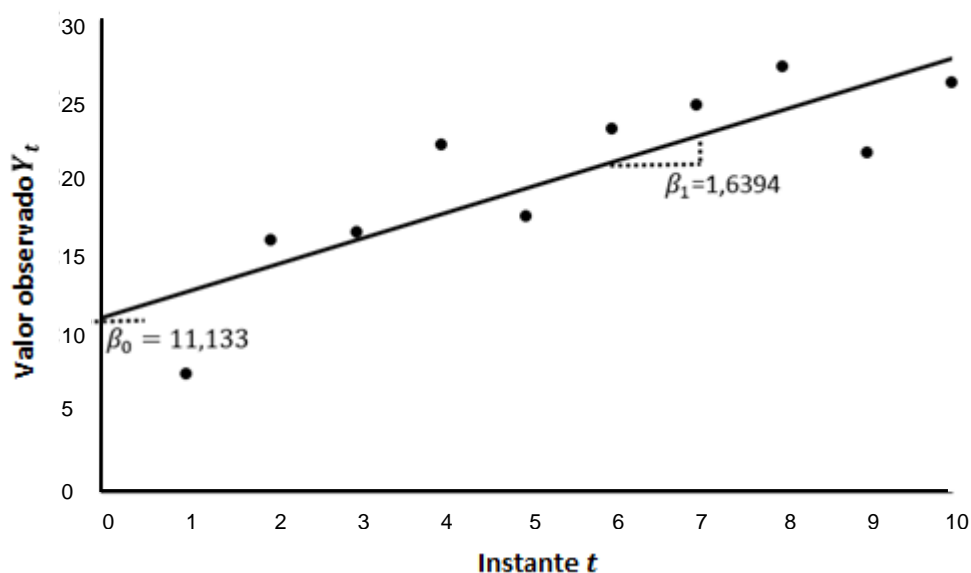


Figura 2.4 – Gráfico de dispersão com reta regressora e parâmetros β_0 e β_1

A reta de regressão representada na Figura 2.4 é obtida por meio da Equação 2.1 sem o elemento residual. O método utilizado para estimar os parâmetros $\hat{\beta}_0$ e $\hat{\beta}_1$ é o método dos mínimos quadrados (MMQ), que tem como objetivo minimizar a diferença dos valores projetados pela reta em relação aos valores observados pela série (HELENE, 2012). Assim, os valores de $\hat{\beta}_0$ e $\hat{\beta}_1$ devem minimizar o valor de Q na Equação 2.2, que representa o somatório do quadrado do residual ε .

$$Q = \sum_{t=1}^n \varepsilon^2 = \sum_{t=1}^n (y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t)^2 \quad (2.2)$$

onde x_t e y_t são os valores observados para as variáveis X e Y no instante t , $\hat{\beta}_0$ e $\hat{\beta}_1$ são os parâmetros que devem ser estimados pelo MMQ e n é o número de dados observados.

Para encontrar o menor valor de Q na Equação 2.2, deve-se derivá-la em relação as variáveis de interesse $\hat{\beta}_0$ e $\hat{\beta}_1$ e igualá-las a zero (PETERNELLI, 2004). Assim, temos que $\hat{\beta}_0$ é estimado conforme apresentado pela Equação 2.3.

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (2.3)$$

onde \bar{x} e \bar{y} são as médias dos valores observados para as variáveis X e Y , e $\hat{\beta}_1$ é parâmetro da função estimado pela Equação 2.4.

$$\hat{\beta}_1 = \frac{\sum_{t=1}^n x_t y_t - \frac{(\sum_{t=1}^n x_t)(\sum_{t=1}^n y_t)}{n}}{\sum_{t=1}^n x_t^2 - \frac{(\sum_{t=1}^n x_t)^2}{n}} \quad (2.4)$$

Uma vez obtidas as estimativas para $\hat{\beta}_0$ e $\hat{\beta}_1$, a reta de regressão que representa o menor valor acumulado dos desvios pode ser escrita como apresentado na Equação 2.1, entretanto, sem o elemento residual.

Uma informação importante para a abordagem de agrupamento apresentada neste trabalho, é o quanto o modelo linear estimado representa o comportamento real dos dados. Assim, o coeficiente de determinação, também conhecido como R^2 , é utilizado para quantificar o quão adequado é o modelo estimado ao ser comparado com a distribuição real dos dados (PETERNELLI, 2004).

Para calcular R^2 , antes é preciso calcular a soma do quadrado da regressão (SQR) e a soma do quadrado total (SQT). A variável SQR é calculada pela Equação 2.5 e indica a diferença entre o valor estimado pelo modelo e a média dos valores observados.

$$SQR = \sum_{t=1}^n (\hat{y}_t - \bar{y})^2 \quad (2.5)$$

onde \hat{y}_t indica o valor estimado pelo modelo, \bar{y} é a média dos valores observados e n o número de observações. A variável SQT , também necessária para o cálculo de R^2 , representa a variação total em Y . A Equação 2.6 apresenta como o cálculo é realizado para a variável.

$$SQT = \sum_{t=1}^n (y_t - \bar{y})^2 \quad (2.6)$$

onde y_t indica o valor observado, \bar{y} a média dos valores observados e n o número de observações. Definidos SQR e SQT , o coeficiente R^2 é calculado como apresentado na Equação 2.7.

$$R^2 = \frac{SQR}{SQT} \quad (2.7)$$

onde $0 \leq R^2 \leq 1$ indica a proporção da variação de Y que é explicada pelo modelo estimado. Se o valor obtido estiver próximo de 1 significa que a reta de regressão representa satisfatoriamente a distribuição real dos dados e, se o valor obtido estiver próximo de 0, o modelo linear estimado não representa adequadamente o comportamento dos dados. A Figura 2.5 apresenta dois exemplos com o valor de R^2 .

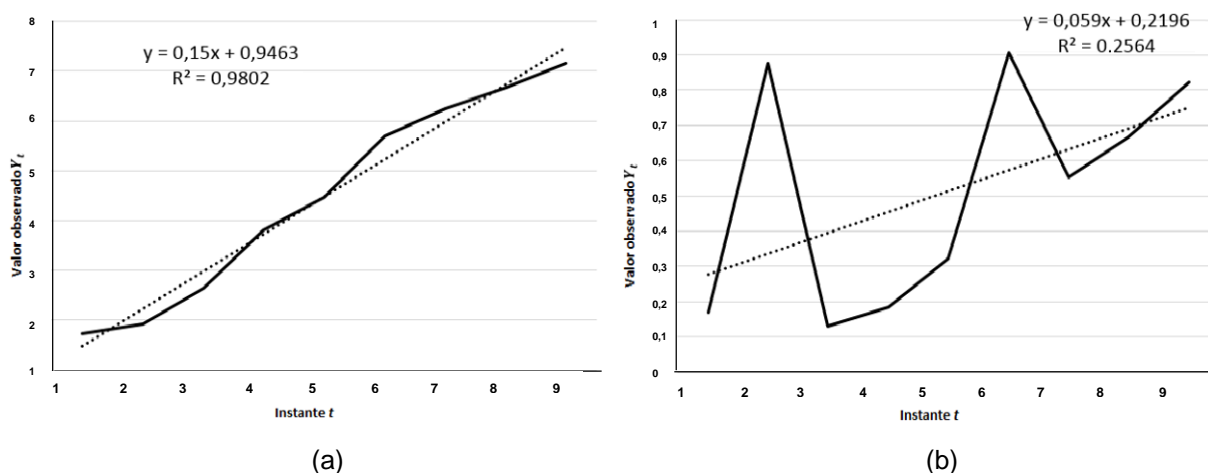


Figura 2.5 – Retas de regressão para a distribuição de dados em duas ST

No exemplo da Figura 2.5(a) é possível observar que a reta de regressão linear gerada pelo modelo matemático estimado, possui um comportamento semelhante ao da distribuição real dos dados, implicando em um coeficiente R^2 próximo a 1. Na Figura 2.5(b) não é possível identificar um padrão comportamental na distribuição dos dados. A equação estimada para a reta de regressão linear na Figura 2.5(b), não representa adequadamente a distribuição real dos dados, implicando em um baixo valor para o coeficiente R^2 .

2.3 Alisamento exponencial simples

Também conhecido como suavização exponencial simples, o alisamento exponencial simples (AES) é considerado um modelo clássico para predição de valores em séries temporais (YAGER, 2013). Seu objetivo é ponderar os valores observados ao longo de uma ST, aumentando ou diminuindo exponencialmente seus pesos para a predição. O modelo matemático apresentado na Equação 2.8 define, de modo recursivo, o valor suavizado para a amostra n .

$$\hat{y}_n = \alpha y_n + (1 - \alpha)\hat{y}_{n-1}, \quad \hat{y}_0 = y_1 \quad (2.8)$$

onde \hat{y}_n é o valor exponencialmente suavizado para a amostra n , que pode ser utilizado como valor de previsão para o instante $n + 1$, y_n é o valor observado no instante n , \hat{y}_{n-1} é o valor suavizado no instante $n - 1$ e, $0 \leq \alpha \leq 1$ é a constante de suavização. Se α for definido próximo de 0, as amostras mais antigas da ST terão maior influência no cálculo do termo \hat{y}_n e, se α for definido próximo a 1 as amostras mais recentes terão maior influência na definição do termo suavizado \hat{y}_n . A Figura 2.6 apresenta um gráfico com os fatores de ponderação dos dados em uma ST com dez amostras para $\alpha = 0,2$, $\alpha = 0,5$ e $\alpha = 0,8$.

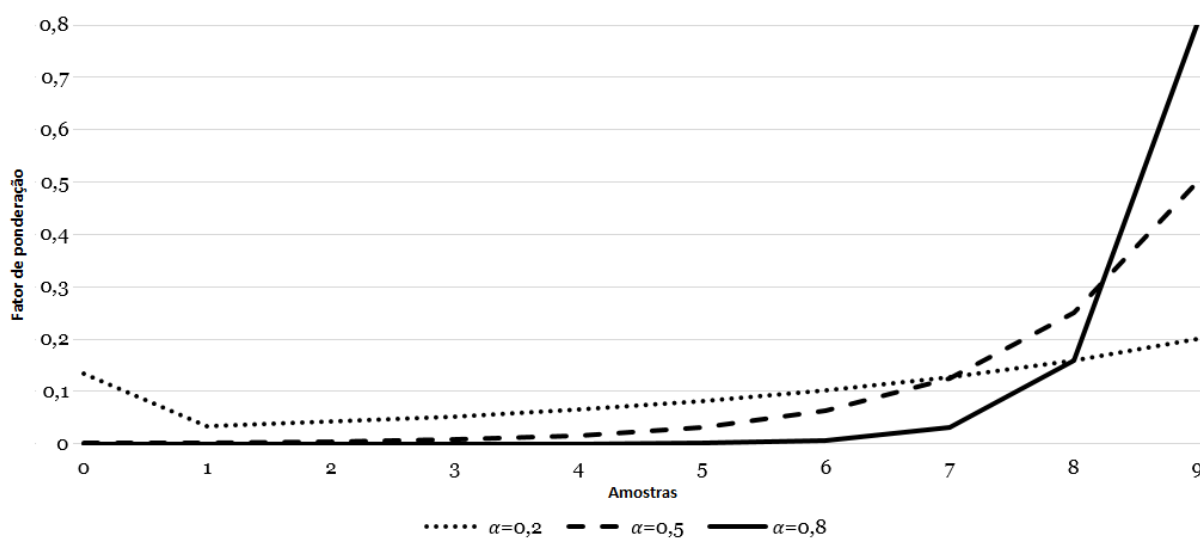


Figura 2.6 – Fatores de ponderação em uma ST com dez amostras para $\alpha = 0,2$, $\alpha = 0,5$ e $\alpha = 0,8$

Ao ser considerada a suavização de uma ST durante todo o seu período de observação, a Equação 2.8 pode ser representada como demonstrado na Equação 2.9.

$$\hat{y}_n = \alpha \sum_{i=0}^{n-1} (1 - \alpha)^i y_{t-i} + (1 - \alpha)^n \hat{y}_0 \quad (2.9)$$

O argumento para o tratamento diferenciado das observações de uma ST, é fundamentado na suposição de que o comportamento mais recente da série tende a conter mais informações sobre o futuro e, portanto, são mais relevantes para a previsão (MORETTIN; TOLOI, 2004). Entretanto, não é possível afirmar que o ajuste de α próximo a 1 é capaz de gerar previsões com melhor acurácia em todos os casos.

2.4 Segmentação de séries temporais

Dentre os principais objetivos para a segmentação de uma ST estão a redução do custo computacional em relação ao processamento, e a análise particionada do comportamento da série (KEOGH, et al., 2004). Basicamente, a segmentação de uma ST é realizada por meio dos algoritmos *top-down*, *bottom-up* e janela deslizante (LI, et al., 2011). Independente de qual modelo matemático será utilizado nos algoritmos, a segmentação pode ser realizada considerando três diferentes parâmetros de restrição na análise (KEOGH, et al., 2004):

- Número máximo de K segmentos ao longo de toda a ST;
- *Threshold* local que avalia os desvios acumulados em um segmento;
- *Threshold* global para controle dos desvios acumulados em todos os segmentos;

Os principais algoritmos de segmentação utilizam a análise dos segmentos de modo independente com a definição de um *threshold* local. A segmentação pelo algoritmo *top-down*, inicialmente considera toda a ST como um só segmento. Para cada possível ponto de divisão, os 2 segmentos resultantes são avaliados. Enquanto a segmentação que gerar o menor erro, ou o maior ganho, estiver satisfazendo o *threshold*, a divisão do segmento é realizada neste respectivo ponto. Devido à necessidade de

cálculo para os 2 segmentos resultantes em cada possível ponto de divisão, esta abordagem é a de maior custo computacional (LI, et al., 2011). Em Park, et al., (1999), uma extensão do algoritmo *top-down* é introduzida. Os autores propõem uma análise para a identificação dos dois pontos extremos da série, ou seja, os pontos onde o menor e o maior erro são gerados com a divisão. A partir destes pontos são criados os primeiros segmentos da série para, na sequência, a abordagem *top-down* clássica ser aplicada normalmente.

O algoritmo *bottom-up* é um complemento natural do *top-down*. Inicialmente são definidos $n/2$ segmentos em uma ST com n dados observados. Para cada novo segmento que possa surgir por meio da combinação de 2 segmentos vizinhos, é feita a análise para verificar a qualidade deste novo segmento. Enquanto o menor erro estiver dentro do *threshold* definido pelo usuário, o processo é repetido.

O algoritmo janela deslizante tem sido amplamente utilizado em ST contínuas (MARIOTE, 2008). O primeiro dado observado à esquerda da ST, ou o primeiro dado disponível, é considerado o início do segmento. Cada novo dado que chega é agregado ao segmento e a análise da qualidade do segmento é realizada. Enquanto o parâmetro de qualidade estiver dentro do *threshold* definido pelo usuário, novos dados são agregados ao segmento. No momento em que um novo dado no instante t é agregado ao segmento e o parâmetro que quantifica a qualidade do segmento exceder o *threshold*, o segmento atual é finalizado na amostra $t - 1$ e um novo segmento tem início com o primeiro dado sendo o valor observado no instante t .

Em Keogh, et al., (2001) foram realizados diversos experimentos considerando os três algoritmos de segmentação. Os autores concluíram que, a partir da análise dos resultados obtidos, a técnica *bottom-up* é a que apresenta os melhores resultados ao ser comparada com as outras duas abordagens.

2.5 Séries Temporais Fuzzy

Muitas vezes, os dados provenientes do mundo real são incompletos, imprecisos ou até mesmo ambíguos. Assim, devido à alta dependência dos valores disponíveis para gerar uma previsão por meio de uma ST, métodos tradicionais podem ter

sua acurácia prejudicada, ocorrência esta amenizada pelo uso dos conjuntos e da lógica fuzzy (LIU, et al., 2009).

Os conceitos da teoria de sistemas fuzzy foram inicialmente aplicados em problemas envolvendo ST por Song e Chissom (1993a, 1993b, 1994), onde os valores observados na ST são representados por conjuntos fuzzy (ZADEH, 1965) dentro do universo de discurso U , onde $U = \{u_1, u_2, \dots, u_n\}$. Um conjunto fuzzy no universo de discurso U pode ser representado como segue:

$$A = f_A(u_1)/u_1 + f_A(u_2)/u_2 + \dots + f_A(u_n)/u_n \quad (2.10)$$

onde f_A é a função de pertinência do conjunto fuzzy A , $f_A: U \rightarrow [0,1]$, $f_A(u_i)$ denota o grau de pertinência de u_i no conjunto fuzzy A e $1 \leq i \leq n$.

Os procedimentos base propostos por Song e Chissom (1993a, 1993b, 1994) para a previsão utilizando STF são: (1) Definir o domínio da variável Y ; (2) Definir os conjuntos fuzzy para a variável Y ; (3) Fuzzificar os dados observados na ST; (4) Construir as RLF a partir da ST fuzzificada; (5) Gerar uma previsão fuzzificada; (6) Defuzzificar a previsão gerada anteriormente; (7) Retornar o valor previsto.

Os procedimentos apresentados acima são a base para a maior parte dos trabalhos com STF (CAI, et al., 2015) (CHEN; CHANG, 2010) (CHATTERJEE, et al., 2011) (SANTOS; CAMARGO, 2014a) (SANTOS; CAMARGO, NO PRELO) e (TANUWIJAYA; CHEN, 2009a). Como descrito, o primeiro passo é a definição do domínio da variável do problema, que é feita com base nas amostras dos dados da variável Y . Após a definição do domínio é necessário determinar quantos conjuntos fuzzy irão representar a variável, bem como a função de pertinência para cada um destes conjuntos.

A fuzzificação a ser realizada no processo 3, consiste em converter cada amostra numérica da ST no conjunto fuzzy de maior pertinência para aquela amostra. Os relacionamentos fuzzy mencionados no item 4, são construídos a partir das amostras fuzzificadas, ou seja, se a amostra no tempo y_{t-1} é representada pelo conjunto fuzzy Alto (A), e a amostra do tempo y_t é representada pelo conjunto fuzzy Médio (M), para $1 < t \leq n$, a relação lógica fuzzy (RLF), entre estas amostras é representada por $A \rightarrow M$. Este procedimento será melhor detalhado ao longo desta seção. O procedimento 5 descrito anteriormente, gera uma previsão fuzzificada, ou seja, o valor previsto é

representado por um conjunto fuzzy. Antes de retornar o valor real ao usuário, o modelo deverá defuzzificar a previsão gerada, que consiste no processo de conversão de um ou mais termos linguísticos em um número real.

De acordo com Song e Chissom (1993a, 1993b, 1994), a previsão de um valor é realizada a partir das RLF extraídas de uma STF. As principais definições relacionadas às STF são apresentadas na sequência (SONG; CHISSOM, 1993a) (SONG; CHISSOM, 1993b) (SONG; CHISSOM, 1994):

Definição 1: Seja $Y(t)$ um subconjunto dos números reais. Considere $f_i(t)$, onde $i = 1, 2, \dots, c$, como um conjunto fuzzy definido no universo de discurso de $Y(t)$. Se $F(t)$ é uma coleção de $f_i(t)$, então $F(t)$ é denominada uma STF de $Y(t)$.

Definição 2: Considere $F(t)$ uma STF e seja $R(t, t - 1)$ uma relação de $F(t)$. Se $R(t, t - 1) = R(t - 1, t - 2)$ para qualquer t , então $F(t)$ é denominada STF invariante no tempo. Caso seja dependente de t , ou seja, $R(t, t - 1)$ pode ser diferente de $R(t - 1, t - 2)$ para algum t , então $F(t)$ é denominada STF variante no tempo.

Definição 3: Considere $F(t)$ uma implicação de $F(t - 1)$, ou seja, $F(t - 1) \rightarrow F(t)$. A relação existente entre $F(t)$ e $F(t - 1)$ é representada por $F(t) = F(t - 1) \circ R(t, t - 1)$, onde " \circ ", originalmente, é definido como o operador Max-Min; e $R(t, t - 1)$ é uma relação fuzzy entre $F(t)$ e $F(t - 1)$. Desta forma $F(t - 1) \rightarrow F(t)$ é denominada relação fuzzy de primeira ordem.

Definição 4: Se $F(t)$ é causado por $F(t - 1), F(t - 2), \dots, F(t - k)$, então esta relação fuzzy é definida como relação de implicação de k -ésima ordem em uma ST, e deve ser representada por $F(t - k), \dots, F(t - 2), F(t - 1) \rightarrow F(t)$.

Para ilustrar o processo de extração das RLF a partir das ST fuzzificadas, considere como exemplo a porcentagem da ocupação média mensal dos leitos de um hotel ao longo de um ano: 0,87; 0,76; 0,52; 0,4; 0,45; 0,22; 0,23; 0,38; 0,44; 0,57; 0,67; 0,88. Supondo que foram definidos os termos linguísticos *Alto*, *Médio* e *Baixo* para esta série, uma possível representação fuzzy da ST é apresentada na Figura 2.7.



Figura 2.7 – Série temporal fuzzificada

Considerando a extração de RLF de primeira ordem, a base de conhecimento definida a partir da ST fuzzificada será composta de onze RLF, conforme apresentado na Figura 2.8.

Alta → Alta	Baixa → Média
Alta → Média	Média → Média
Média → Média	Média → Média
Média → Média	Média → Média
Média → Baixa	Média → Média
Baixa → Baixa	Média → Alta

Figura 2.8 – RLF extraídas da série temporal fuzzificada

As RLF devem ser extraídas a partir da primeira amostra da STF. Assim, ao analisarmos a ST acima, temos a primeira RLF definida como *Alta* → *Alta*, na sequência *Alta* → *Média* e assim por diante. Desta forma, a base de conhecimento será composta por $n - k$ RLF, onde n é o número de amostras observadas na série e k indica a ordem das RLF como descrito na Definição 4. Na Figura 2.9 é apresentada uma base de conhecimento composta por RLF de 3ª – ordem.

<i>A, A, M</i> → <i>Média</i>	<i>B, B, M</i> → <i>Média</i>
<i>A, M, M</i> → <i>Média</i>	<i>B, M, M</i> → <i>Média</i>
<i>M, M, M</i> → <i>Baixa</i>	<i>M, M, M</i> → <i>Média</i>
<i>M, M, B</i> → <i>Baixa</i>	<i>M, M, M</i> → <i>Alta</i>
<i>M, B, B</i> → <i>Média</i>	

Figura 2.9 – RLF de 3ª – ordem

Diferente do que ocorre nas regras fuzzy convencionais, mesmo com antecedentes iguais e consequentes diferentes, como ocorre nas RLF com os antecedentes *M, M, M*, esta situação não é considerada um conflito, pois representa o real comportamento da série ao longo do período de observação. O valor a ser apresentado como previsão, deve ser calculado pelo método de defuzzificação definido pelo modelo.

2.6 Pré-processamento em séries temporais fuzzy

Nesta seção serão apresentados alguns dos conceitos relacionados ao pré-processamento dos dados em ST, visando a identificação de *outliers* e a definição da quantidade e do suporte dos conjuntos fuzzy em uma variável.

No processo de identificação e remoção dos *outliers* apresentado em (BARNETT; LEWIS, 1994), o primeiro passo é definir a matriz quadrada H , como apresentado na Equação 2.11.

$$H = z(z^T z)^{-1} z^T \quad (2.11)$$

onde z^T é a transposta da matriz apresentada na Equação 2.12,

$$z = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \quad (2.12)$$

Definida a matriz H , o próximo passo é calcular, a partir da Equação 2.13, o índice *Residual Student* para cada amostra da ST.

$$RStudent_t = \frac{e_t}{\hat{\sigma}^{(t)} \sqrt{1 - h_t}} \quad (2.13)$$

Na equação 2.13, $\hat{\sigma}^{(t)}$ representa o desvio padrão sem o elemento t da série, h_t é o t – ésimo elemento diagonal da matriz H e e_t é definido pela Equação 2.14, onde y_t representa a amostra da ST para a qual o *Residual Student* é calculado e y_j é o dado observado no instante j . Conforme sugerido por Liu et. al. (2009), se $RStudent_t > 2,5$, então y_t pode ser considerado um *outlier* (SANTOS; CAMARGO, 2013).

$$e_t = y_t - \frac{\sum_{j=1}^n y_j}{n} \quad (2.14)$$

Em (TANUWIJAYA; CHEN, 2009b) é apresentado um método para a definição da quantidade de termos linguísticos que representarão os dados observados em uma ST. O procedimento é apresentado na sequência:

- **Passo 1.** Ordenar a ST em uma sequência numérica ascendente $y_1, y_2, \dots, y_t, \dots, y_n$, onde y_1 é o menor valor da ST e y_n o maior dado numérico observado, e $1 \leq t \leq n$.
- **Passo 2.** Calcular o valor τ de parada do algoritmo conforme a Equação 2.15.

$$\tau = \frac{\sum_{t=1}^{n-1} (y_{t+1} - y_t)}{n - 1} \quad (2.15)$$

- **Passo 3.** Inicialmente, considerar cada valor da ST como um *cluster* de um só elemento.
- **Passo 4.** Considere a existência de c *clusters*. Calcule os centros dos *clusters* conforme apresentado na Equação 2.16, onde p é o número de elementos no $cluster_k$, y_j são os valores numéricos dos elementos pertencentes ao $cluster_k$, e $1 \leq k \leq c$.

$$centro_cluster_k = \frac{\sum_{j=1}^p y_j}{m} \quad k = 1, 2, \dots, c. \quad (2.16)$$

- **Passo 5.** Calcular a distância entre o *cluster* k e $k + 1$, onde $k = 1, 2, \dots, c - 1$, conforme Equação 2.17.

$$distância_{k,k+1} = |centro_cluster_k - centro_cluster_{k+1}| \quad (2.17)$$

- **Passo 6.** Se a menor distância entre os *clusters* for inferior a τ , faça a união dos *clusters* mais próximos e retorne ao Passo 4. Caso contrário, encerre o algoritmo.

Encerrado o procedimento para definir a quantidade de termos linguísticos, os valores que representarão o parâmetro central dos conjuntos fuzzy são definidos pela Equação 2.16.

2.7 Considerações finais

Neste capítulo foram apresentados os conceitos de séries temporais clássicas e séries temporais fuzzy. As técnicas de alisamento exponencial simples, regressão linear e de segmentação, serão utilizadas como ferramentas de apoio nas abordagens apresentadas nesta tese para a predição de valores e para a análise comportamental das séries. O modelo de pré-processamento desenvolvido utiliza como base os conceitos apresentados na Seção 2.6.

No Capítulo 3 serão apresentados os conceitos referentes à análise comportamental das ST, com foco especial para as métricas e para os algoritmos de agrupamento que são utilizados na análise das séries.

Capítulo 3

AGRUPAMENTO DE SÉRIES TEMPORAIS

Neste capítulo serão apresentados os conceitos associados ao agrupamento de séries temporais. Um algoritmo de agrupamento, ou de clusterização, tem como objetivo agrupar séries temporais com comportamentos semelhantes em um mesmo *cluster* e manter as séries temporais distintas em *clusters* diferentes. Uma das principais vantagens do agrupamento fuzzy, quando comparado ao agrupamento tradicional, é o fato dos elementos de dados possuírem graus de pertinência em cada um dos *clusters* identificados. Dentro do contexto deste trabalho, cada ST avaliada é considerada um elemento de dados.

Além da escolha do algoritmo para o agrupamento das ST, outra importante característica a ser definida é qual métrica será utilizada para quantificar o quão similar ou dissimilar são duas séries.

Segundo Theodoridis e Koutroumbas (2009), em alguns casos é necessário verificar se o modo como os dados estão distribuídos não inviabiliza a definição dos

grupos. Confirmada a tendência de agrupamento dos dados, ainda assim os resultados podem não ser os esperados. Para ilustrar, considere o exemplo apresentado na Figura 3.1.

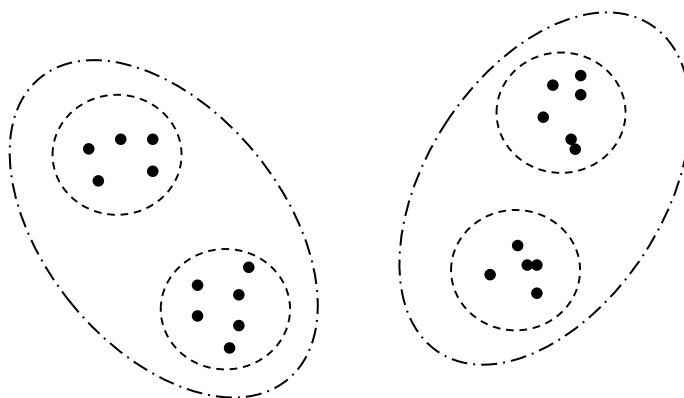


Figura 3.1 - Possíveis cenários para o agrupamento de dados.
(Adaptada de (THEODORIDIS; KOUTROUMBAS, 2009))

Uma possível pergunta a respeito do cenário apresentado na Figura 3.1 é quantos *clusters* existem na distribuição dos dados? Alguns poderiam dizer 2. Outros acreditam que sejam 4, e ainda há quem poderia dizer 3. Em casos como este, muitas vezes são realizados experimentos com diferentes números de *clusters*, e os agrupamentos obtidos são analisados para identificar a quantidade ideal de grupos.

Os métodos para agrupamento de dados podem ser não supervisionados ou semissupervisionados. Nos métodos não supervisionados, não existem dados rotulados. Desta forma, o algoritmo deve identificar os dados com estrutura semelhante e agrupá-los em um mesmo *cluster*, sem o auxílio dos rótulos predefinidos. Os métodos semissupervisionados possuem como característica a rotulação de apenas alguns dados disponíveis para treinamento. A partir dos dados rotulados, o algoritmo identifica os padrões na distribuição dos dados e, conseqüente, os *clusters* presentes na base de dados.

O agrupamento de séries temporais tem como característica considerar cada série como um elemento de dado, ou seja, uma ST com 100 valores observados, por exemplo, é considerada um elemento único com 100 amostras. Na sequência serão apresentadas algumas das principais métricas para agrupamento de séries temporais, além dos algoritmos base utilizados no agrupamento das mesmas.

3.1 Métricas para séries temporais

Um ponto fundamental no processo de agrupamento é a escolha da métrica a ser utilizada na análise das séries temporais. Dentre as características que devem ser observadas, o tipo e o comportamento dos dados são de fundamental importância na definição da métrica. Assim, nesta seção são apresentadas algumas das principais métricas utilizadas no agrupamento de séries temporais.

3.1.1 Distância Euclidiana

Assim como no agrupamento de dados convencionais, a distância euclidiana também é uma métrica utilizada para o agrupamento de ST (DEVIJVER; KITTLER, 1982). Sejam X_t e Y_t dois vetores que representam os valores reais observados em duas séries temporais. A dissimilaridade entre as duas séries é calculada pela Equação 3.1 (DEZA; DEZA, 2009).

$$d_E(X_t, Y_t) = \sqrt{\sum_{t=1}^n (x_t - y_t)^2} \quad (3.1)$$

Como é possível observar, para calcular a dissimilaridade utilizando a distância euclidiana é necessário que ambas as ST possuam a mesma quantidade n de amostras. Alguns dos problemas em utilizar a distância euclidiana como métrica para analisar séries temporais, são a alta sensibilidade a ruídos e as pequenas defasagens no tempo em relação ao comportamento das séries (AGRAWAL, et al., 1995). Como exemplo, considere as ST ilustradas na Figura 3.2.

O resultado apresentado pela distância euclidiana ao analisar a similaridade entre cada uma das duas ST e a linha tracejada, diz que há maior similaridade entre os valores de cada uma das duas ST quando comparadas com a linha tracejada, do que quando comparadas a Série 1 com a Série 2, que realmente possuem semelhanças na prática.

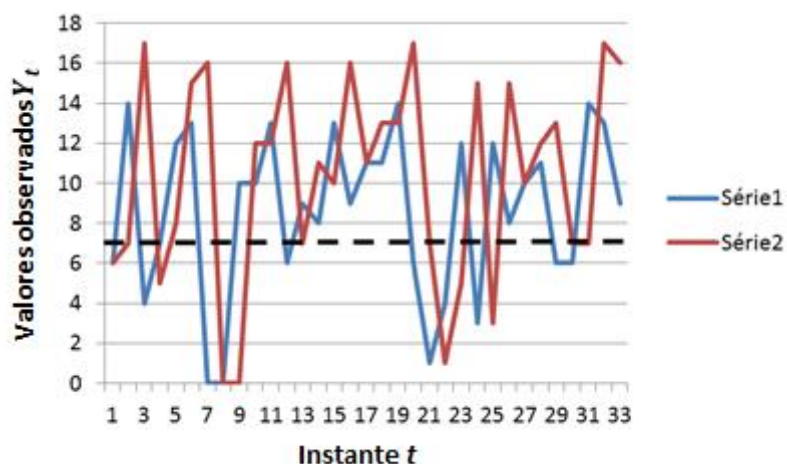


Figura 3.2 – Distância Euclidiana. Uma pequena variação de fase pode ocasionar uma distorção do resultado esperado

3.1.2 Distância de Hamming

Outra métrica baseada na divergência entre os pontos é a distância de Hamming, que nada mais é do que a soma das amostras que diferem nos vetores X_t e Y_t no mesmo tempo t . O cálculo é realizado conforme apresentado na Equação 3.2 (HAMMING, 1950).

$$d_H(X_t, Y_t) = \sum_{t=1}^n h(x_t, y_t) \quad (3.2)$$

Quando $x_t = y_t$ então $h(x_t, y_t) = 0$, senão $h(x_t, y_t) = 1$. Quanto menor o valor de d_H , mais semelhantes são as ST de acordo com a métrica. Se necessário, pode ser definida uma margem de aceitação para que as amostras sejam consideradas iguais.

3.1.3 Coeficiente de correlação de Pearson

Como alternativa as métricas anteriores, a similaridade entre duas ST representadas por X_t e Y_t pode ser calculada por meio da correlação de Pearson, representada na Equação 3.3 (DEZA; DEZA, 2009).

$$cc(X_t, Y_t) = \frac{\sum_{t=1}^n (x_t - \bar{x}_t)(y_t - \bar{y}_t)}{S_{X_t} S_{Y_t}} \quad (3.3)$$

onde \bar{y}_t é a média dos valores da série temporal Y_t , y_t é o valor observado no instante t , e S_{Y_t} representa o desvio padrão da ST calculado pela Equação 3.4.

$$S_{X_t} = \left[\sum_{t=1}^n (y_t - \bar{y}_t) \right]^{1/2} \quad (3.4)$$

O coeficiente de correlação de Pearson cc tem seu valor definido entre -1 (menos um) e 1 (um). Se as ST apresentarem tendências similares, o resultado será próximo de 1. Se forem tendências inversamente proporcionais, o resultado será próximo de -1. Se o resultado apresentado for 0, a análise indica que não existe relação linear entre as ST analisada. Uma análise detalhada desta métrica pode ser encontrada em (FIGUEIREDO; SILVA, 2009)

3.1.4 Short Time Series

Outra medida de dissimilaridade para números reais é a *Short Time Series* (STS), apresentada na Equação 3.5. Considerando cada ST como uma função linear por partes, a métrica tem como base de cálculo a soma do quadrado das diferenças entre as amostras das ST. Esta métrica considera que a dissimilaridade entre as ST pode ser escrita como um segmento de reta. Desta forma, a métrica analisa a inclinação destas retas para definir sua dissimilaridade (MOLLER-LEVET, et al., 2003).

$$d_{STS}(X_t, Y_t) = \sqrt{\sum_{t=1}^{n-1} \left(\frac{y_{t+1} - y_t}{(t+1) - t} - \frac{x_{t+1} - x_t}{(t+1) - t} \right)^2} \quad (3.5)$$

Da mesma forma que nas equações anteriores, X_t e Y_t são dois vetores que representam os valores observados de duas ST. A posição de uma amostra x_t ou y_t na série temporal é denotada por t e n é o número total de amostras.

3.1.5 Similaridade triangular ou similaridade do cosseno

A similaridade triangular entre duas ST é calculada considerando o cosseno entre os dois vetores que as representam. O primeiro passo para o cálculo é a normalização das amostras das ST X_t e Y_t , como apresentado na Equação 3.6, onde \hat{x}_t representa o valor normalizado no instante t (DEZA; DEZA, 2009).

$$\hat{x}_t = \frac{x_t}{(\sum_{k=1}^n x_k^2)^{1/2}}, \quad t = 1, 2, \dots, n. \quad (3.6)$$

Normalizadas todas as amostras de ambas as ST, a similaridade do cosseno é definida pela Equação 3.7.

$$d_T(X_t, Y_t) = \frac{\sum_{t=1}^n x_t y_t}{(\sum_{t=1}^n x_t^2)^{1/2} (\sum_{t=1}^n y_t^2)^{1/2}} = \sum_{t=1}^n \hat{x}_t \hat{y}_t \quad (3.7)$$

O valor $d_T(X_t, Y_t)$ é definido entre 1 (um) e -1 (menos um). Quando os dois vetores que representam as ST estão sobrepostos e com mesma direção, ou seja, são totalmente similares, o resultado da equação é 1. Quanto mais próximo de -1 for o resultado, menos similares são os vetores.

Um problema identificado durante alguns experimentos realizados com esta métrica, é a não relevância das grandezas escalares na análise. Ao comparar a ST $X_t = \{3, 6, 7, 5\}$ com a ST $Y_t = \{3000, 6000, 7000, 5000\}$, a métrica informa que as ST são totalmente semelhantes, retornando o valor um. Desta forma, em alguns contextos como o financeiro, por exemplo, este pode ser um problema ao optar pela similaridade triangular.

3.1.6 *Dynamic Time Warping*

Dynamic Time Warping (DTW) (BELLMAN, 1957) é um algoritmo amplamente utilizado para calcular a dissimilaridade entre duas ST e que tem apresentado os resultados mais significativos para o agrupamento de séries temporais (XI, et al., 2006). Além da flexibilidade na análise comparativa entre duas séries, o algoritmo DTW também apresenta como vantagem a não necessidade de que ambas as séries tenham o

mesmo número n de amostras. Recentemente, diversas propostas de extensão ao modelo original tem sido apresentadas (GÓRECKIA; ŁUCZAK, 2015) (IZAKIANA, et al., 2015) (JEONG, et al., 2011) (KIM, et al., 2014) (LI, et al., 2014) e (SU, et al., 2014).

Ao contrário do que ocorre na distância Euclidiana e na distância de Hamming, o modelo DTW realiza uma análise comparativa não linear das ST, como ilustrado na Figura 3.3.

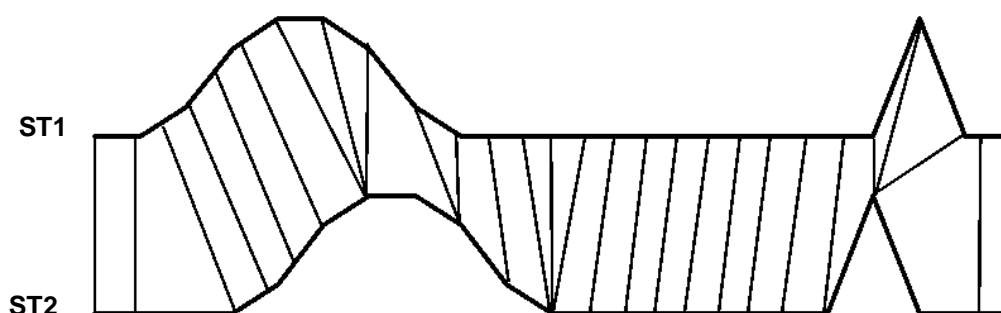


Figura 3.3 – Duas séries temporais analisadas pelo algoritmo DTW

O modelo baseia-se em um algoritmo que tem como objetivo encontrar a menor dissimilaridade entre duas ST. Considere $X_t = \{x_1, x_2, \dots, x_m\}$ uma ST de tamanho m e $Y_t = \{y_1, y_2, \dots, y_n\}$ uma ST de tamanho n . Para calcular a dissimilaridade por meio do algoritmo DTW, o primeiro passo é definir uma matriz $n \times m$ como ilustrado na Figura 3.4, onde a célula (i, j) representa a distância euclidiana $d_E(x_i, y_j)$ entre as amostras x_i e y_j .

1o.		ST x_t				
		1	2	3	4	5
ST y_t	1	0,05	0,04	0,08	0,11	0,12
	2	0,14	0,13	0,17	0,2	0,21
	3	0,24	0,23	0,27	0,3	0,31
	4	0,37	0,36	0,4	0,43	0,44
	5	0,54	0,53	0,57	0,6	0,61

Figura 3.4 – Matriz de distância entre os valores observados em duas ST com 5 amostras cada uma

O segundo passo é definir os valores acumulados para a linha 1 e para a coluna 1. A partir da definição destes valores, cada célula restante da matriz deve ser preenchida considerando o seu próprio valor, somado com a menor distância acumulada de seus vizinhos imediatos. A Figura 3.5 ilustra o processo.

2o.		ST X				
		1	2	3 t	4	5
ST _t Y	1	0,05	0,09	0,17	0,28	0,4
	2	0,19	0,18			
	3	0,43				
	4	0,8				
	5	1,34				

Figura 3.5 – Preenchimento da matriz DTW

A partir do exemplo apresentado na Figura 3.4, as células da linha 1 da matriz DTW na Figura 3.5 foram preenchidas como apresentado na Equação 3.8.

$$DTW(1, j) = \sum_{c=1}^j (1, c) , \quad j = 1, 2, \dots m. \tag{3.8}$$

onde m é o número de colunas da matriz, e $DTW(1, j)$ é a distância acumulada calculada para cada célula da primeira linha, considerando como base os valores da matriz na Figura 3.4. De modo análogo, as células da primeira coluna da matriz DTW na Figura 3.5, também é preenchida a partir do valor acumulado das células anteriores da coluna 1.

A partir da célula (2,2), em destaque na Figura 3.5, o preenchimento é realizado buscando o menor custo acumulado dos vizinhos imediatos. A célula (2,2) possui o valor 0,18 pois ela tem seu valor 0,13 (Figura 3.4) e, seu vizinho imediato de menor custo acumulado na Figura 3.5 é a célula (1,1), com valor 0,5, totalizando o valor 0,18. O mesmo procedimento é aplicado às demais células. A Figura 3.6 apresenta a matriz final do exemplo utilizado, onde o valor da célula (n, m) indica a dissimilaridade obtida pelo algoritmo DTW entre as séries temporais X_t e Y_t .

$n \times m$ - ésima		ST X _t				
		1	2	3	4	5
ST _t Y	1	0,05	0,09	0,17	0,28	0,4
	2	0,19	0,18	0,26	0,37	0,49
	3	0,43	0,41	0,45	0,56	0,68
	4	0,8	0,77	0,81	0,88	1
	5	1,34	1,3	1,34	1,41	1,49

Figura 3.6 – Matriz DTW com o cálculo final da dissimilaridade entre duas ST

Definida a matriz da Figura 3.6, o algoritmo DTW inicia o processo recursivo para identificar qual o caminho gerou o valor da célula (n, m) . Em um processo reverso ao utilizado para preencher a matriz, a partir da posição (n, m) , o algoritmo avalia qual a célula vizinha de menor custo considerando as regras apresentadas na Equação 3.9.

$$W = \begin{cases} (1, j - 1), & \text{se } i = 1 \\ (i - 1, 1), & \text{se } j = 1 \\ \min [d(i - 1, j), d(i, j - 1), d(i - 1, j - 1)], & \text{caso contrário} \end{cases} \quad (3.9)$$

onde i e j indicam os índices das células atuais. Identificada a célula vizinha de (n, m) com menor custo, o algoritmo atualiza os índices i e j da célula atual e avalia novamente, a partir desta nova célula, qual a nova célula vizinha de menor custo de acordo com as restrições da Equação 3.9. O processo é repetido até que o algoritmo chegue à célula $(1,1)$, definindo assim o caminho $W = w_1, w_2, w_3, \dots, w_k$ de menor custo, onde $w_1 = (1,1)$ e $w_k = (n, m)$.

No exemplo apresentado na Figura 3.6, o caminho de menor custo é definido pela diagonal principal. Entretanto, cabe ressaltar que o caminho de menor custo pode passar por qualquer célula dentro da matriz, o que pode acabar gerando uma distorção na análise das séries. Considere o exemplo apresentado na Figura 3.7.

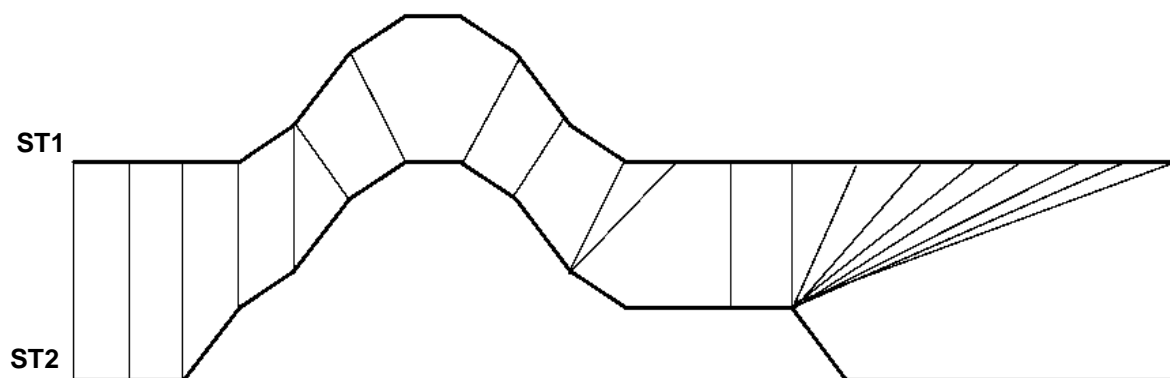
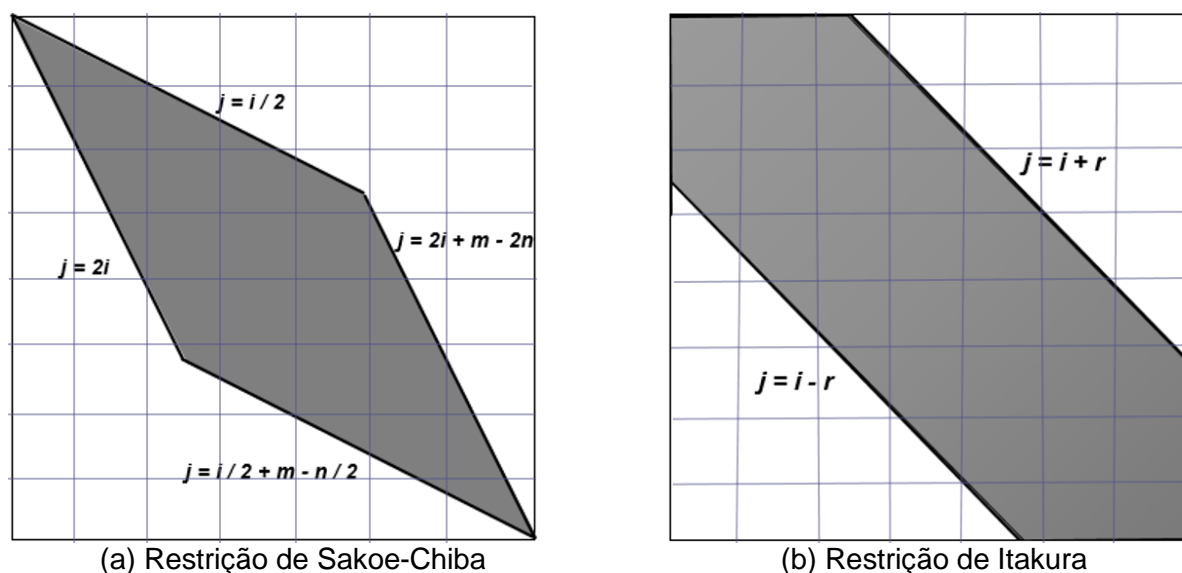


Figura 3.7 – Caminho DTW com distorção na análise

A Figura 3.7 ilustra uma distorção na análise onde amostras próximas à metade do período de observação de uma ST são comparadas com as amostras observadas

no final de outra ST, o que pode conduzir a um resultado que não representa adequadamente a comparação entre as séries. Para evitar situações como esta, existem duas abordagens principais de restrição.

A restrição de Sakoe-Chiba (SAKOE; CHIBA, 1978) limita o espaço de preenchimento e busca na matriz DTW a partir da diagonal principal, onde o usuário define um parâmetro r que indica o tamanho da restrição, ou seja, o tamanho da área de busca. Outra restrição amplamente utilizada é o paralelograma de Itakura (ITAKURA, 1975), que define a área de busca por meio de um paralelograma na matriz. Ambas as restrições e o modo de cálculo são apresentadas na Figura 3.8.



(a) Restrição de Sakoe-Chiba

(b) Restrição de Itakura

Figura 3.8 – Restrições de Sakoe-Chiba e de Itakura (Adaptada de (HENNIGER; MULLER, 2007))

Na restrição de Sakoe-Chiba, apresentada na Figura 3.8(a), o preenchimento da matriz DTW e a área de busca pelo caminho W de menor custo, são definidos de acordo com o parâmetro r tendo como referência a linha i . Em uma matriz quadrada, por exemplo, quando temos $r = 0$, a área de busca é limitada a diagonal principal, pois $j = i$. Na restrição de Itakura, apresentada na Figura 3.8(b), o tamanho das ST é o parâmetro de restrição da área de busca e do preenchimento da matriz DTW para a definição do caminho W .

3.2 Agrupamento de séries temporais

Assim como ocorre no agrupamento de dados convencional, a escolha do algoritmo de agrupamento de ST deve considerar alguns aspectos como, por exemplo, se os dados analisados são contínuos ou discretos, e se as séries temporais possuem o mesmo tamanho ou não.

Algumas abordagens realizam o pré-processamento das ST antes da execução do algoritmo de agrupamento. Em outros trabalhos o agrupamento é realizado a partir dos dados brutos. A primeira abordagem, que envolve o pré-processamento nos dados, é capaz de reduzir a dimensão dos dados eliminando algumas amostras, atributos ou ainda extraíndo parâmetros para a análise das séries, reduzindo assim, consideravelmente o custo computacional necessário para processamento (YEH, et al., 2007). Dentre outros métodos para manipulação dos dados, podemos citar a transformação discreta de Fourier (AGRAWAL, et al., 1993) e (JANACEK, et al., 2005) e a poda (BAGNALL, et al., 2006) e (BAGNALL; JANACEK, 2005).

A segunda abordagem para agrupamento das ST utiliza os dados brutos, sem nenhum tipo de transformação ou poda dos dados. A principal característica destes métodos de agrupamento é a métrica utilizada na análise, uma vez que os dados são analisados da forma como foram disponibilizados (LIAO, 2005).

Duas linhas de pesquisas se desenvolveram para o agrupamento de ST com o auxílio dos algoritmos convencionais. Na primeira delas são realizadas alterações nas métricas utilizadas pelos algoritmos, preservando as demais características das propostas originais. Na segunda linha de pesquisa os dados das ST são manipulados para, assim, serem tratados como dados comuns em um algoritmo de agrupamento convencional. Neste trabalho é utilizada a primeira abordagem associada às adequações nos algoritmos de agrupamento.

Na sequência são apresentados alguns dos algoritmos mais utilizados para agrupamento de séries temporais, como o do Vizinheiro Mais Próximo, K-means e Fuzzy C-means.

3.2.1 K-means

Um dos mais simples e utilizados algoritmos de agrupamento de dados é o *K-means* (MACQUEEN, 1967). O objetivo do *K-means*, como em qualquer outro algoritmo de agrupamento, é agrupar os dados similares em um mesmo *cluster* e deixar os dados dissimilares em *clusters* diferentes. Isto é feito utilizando uma métrica como parâmetro de análise, onde o objetivo é minimizar a distância média dos elementos de um *cluster* para o seu centroide.

Vários são os trabalhos baseados no algoritmo *K-means*. Alguns destes trabalhos focam na otimização dos resultados por meio de ajustes nos vetores de inicialização dos *clusters* (REDDY; PRASANTA, 2012) e (XIE, et al., 2014) e outras extensões do *K-means* trabalham com o agrupamento de séries temporais com foco na métrica utilizada (IZAKIANA, et al., 2015) (NIEEATTRAKUL; RATANAMAHATANA, 2007) e (THINH; ANH, 2012).

Em sua versão original, a quantidade de k *clusters* e os centroides iniciais que vão representa-los são definidos aleatoriamente. Definidos os centroides iniciais, o algoritmo então deve calcular a distância euclidiana de cada um dos elementos para os k centroides. Cada dado deve ser associado ao centroide mais próximo. Realizada a primeira iteração, o algoritmo atualiza os valores dos k centroides por meio da média dos elementos pertencentes a cada *cluster*. Na sequência é feito um novo cálculo da distância entre todos os dados e os novos centroides para, então, serem atualizados os elementos de cada *cluster*. Esta iteração ocorre até o algoritmo convergir ou por uma quantidade n de iterações definida pelo usuário.

Em geral, o algoritmo *k-means* apresenta bons resultados, mas é sensível aos centroides iniciais, ou seja, de acordo com os centros de *clusters* definidos aleatoriamente, os resultados finais apresentados pelo algoritmo podem ser diferentes e, portanto, não apresentar o agrupamento esperado. Quando os dados são conhecidos, uma possível solução para garantir um resultado ao menos próximo do esperado, é a definição dos centroides iniciais pelo usuário. A Figura 3.9 representa duas situações de inicializações diferentes dos centros de *cluster* com o mesmo conjunto de dados.

Na Figura 3.9 foram definidas 32 (trinta e duas) amostras de dados aleatórios e 5 (cinco) centros de *clusters*. As Figuras 3.9 (A) e 3.9 (B) apresentam duas configurações iniciais diferentes para os cinco centros de *clusters*, representados pelos quadrados. Os *clusters* obtidos pela configuração inicial apresentada em 3.9 (A), apresentados no quadro (C), visualmente representam melhor a distribuição dos dados que os *clusters* apresentados no quadro (D), gerados pela configuração inicial apresentada em 3.9 (B). Esta divergência ocorre em razão dos diferentes valores iniciais para os centros de *cluster*. Em ambos os casos, o algoritmo convergiu após 3 iterações.

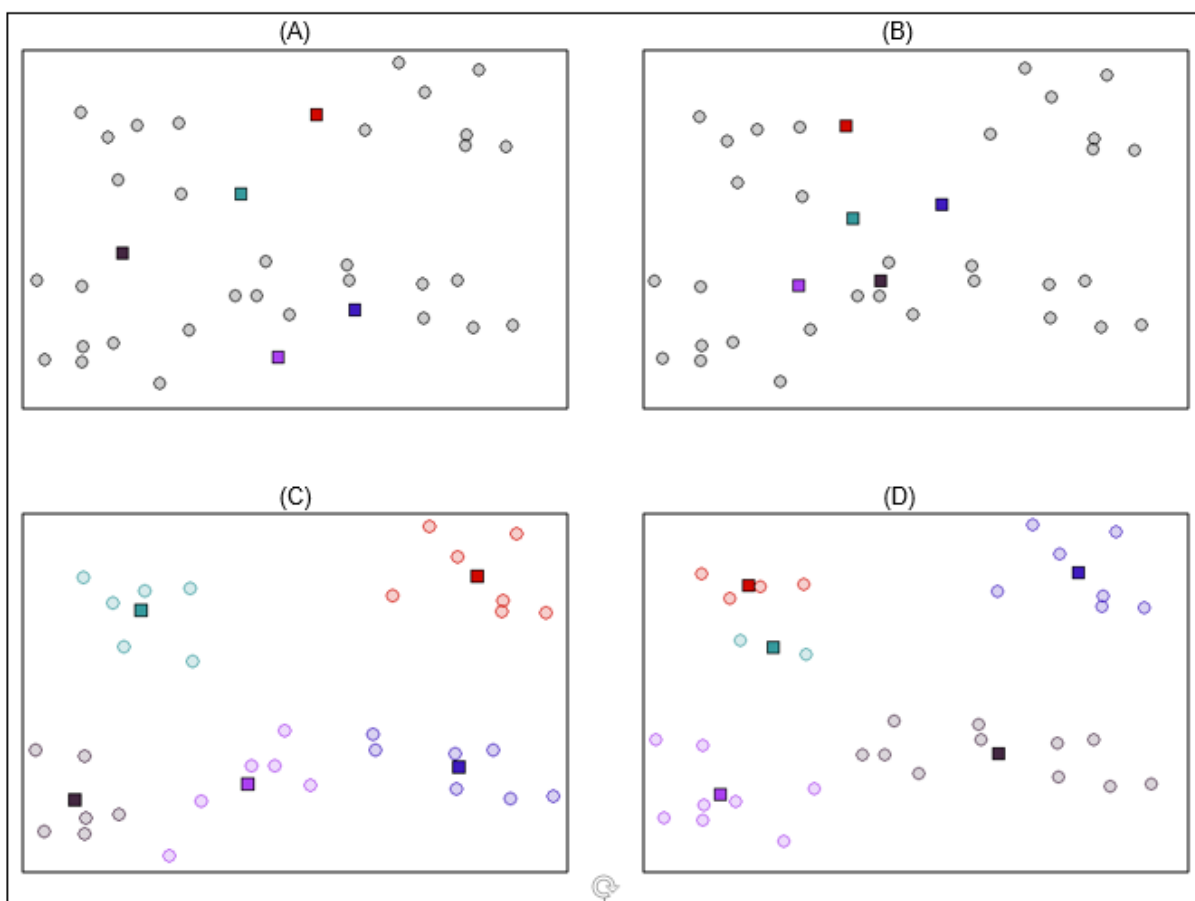


Figura 3.9 – Diferentes centróides obtidos pelo *k-means*

Outro ponto que requer cuidado ao escolher o *K-means* como algoritmo de agrupamento é a presença de *outliers* nos dados. Em razão do centro de *cluster* ser calculado com base na média dos dados de cada grupo, *outliers* tendem a desviar o centro do *cluster* em sua direção, principalmente em *clusters* com poucos elementos.

Desta forma, torna-se extremamente importante o tratamento dos dados para detecção e eliminação de *outliers* antes da execução do algoritmo.

3.2.2 Fuzzy C-Means

O algoritmo de agrupamento *K-means* tem como característica gerar partições de modo que cada elemento pertença somente a um *cluster*. Por sua vez, o algoritmo de agrupamento Fuzzy C-Means (BEZDEK, 1981) é uma extensão do agrupamento clássico, e tem como característica a associação de um grau de pertinência dos elementos de dados a cada um dos *clusters* identificados.

Seja $X = \{x_1, x_2, x_3, \dots, x_n\}$, onde n representa o número de elementos do conjunto de dados X . Dado um número c de *clusters*, uma pseudo-partição fuzzy $\wp = \{A_1, A_2, \dots, A_c\}$ deve satisfazer a restrição apresentada na Equação 3.10, onde $A_i(x_k)$ é o grau de pertinência de (x_k) na pseudo-partição A_i .

$$\sum_{i=1}^c A_i(x_k) = 1 \quad (3.10)$$

Em um agrupamento fuzzy, um *cluster* i possui a restrição de não poder ter os n elementos do conjunto X com o grau de pertinência igual a um (máximo) e nem com o grau de pertinência igual a zero (*cluster* vazio). Para isso, a Equação 3.11 deve ser satisfeita.

$$0 < \sum_{i=1}^c A_i(x_k) < n \quad (3.11)$$

Considerando os protótipos iniciais do algoritmo, o melhor agrupamento do Fuzzy C-Means é obtido por meio da minimização da Equação 3.12.

$$J_q(U_h, C) = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m d(x_k, c_i) \quad (3.12)$$

Na Equação 3.12, U_h é a matriz de pertinência com dimensão $c \times n$, onde c é a quantidade de *clusters* e n o total de dados na base X . Desta forma, u_{ik} representa

a pertinência do elemento k no cluster i , m é o índice de fuzzificação e $d(c_i, x_k)$ é a distância entre o cluster c_i e o elemento x_k . A Equação 3.12 ainda deve respeitar a restrição apresentada na Equação 3.10. O cálculo utilizado para a atualização dos centróides é apresentado na Equação 3.13.

$$c_i = \frac{\sum_{k=1}^n u_{ik} x_k}{\sum_{k=1}^n u_{ik}} \quad (3.13)$$

onde c_i é o valor do centroide, n é o número de elementos e u_{ik} é o grau de pertinência do elemento x_k no cluster i . O algoritmo deverá ser executado até que ocorra um número máximo de iterações ou até que não haja mais ganho significativo nos novos agrupamentos obtidos após ser finalizada uma iteração. Cada célula da matriz de pertinência U_h deve ser atualizada pela Equação 3.14.

$$A_i(x_k) = \left[\sum_{j=1}^c \left(\frac{\|x_k - c_i\|^2}{\|x_k - c_j\|^2} \right)^{\frac{1}{m-1}} \right]^{-1} \quad (3.14)$$

Na Equação 3.14, $A_i(x_k)$ é o grau de pertinência do elemento x_k do cluster c_i , c é o número total de clusters, $m \geq 0$ é o nível de fuzzificação e se $m = 1$ os graus de pertinência são definidos como 0 ou 1. Uma visão geral do Fuzzy C-means é apresentada no Algoritmo 3.2.

Algoritmo 3.2: Fuzzy C-Means

1. Definir a quantidade c de clusters;
 2. Definir o número máximo de iterações (γ) e o ganho mínimo necessário δ para que o algoritmo continue sendo executado;
 3. Inicializar aleatoriamente a matriz U_h de pertinências;
 4. Atribuir $cont = 0$, onde $cont$ é contador de iterações;
 5. repita
 6. $cont \leftarrow cont + 1$;
 7. Atualizar $U_h (J_q(U_h, C))$;
 8. Atualizar centros de cluster C_i ;
 9. até que $\|c_i^y - c_i^{y+1}\| < \delta$ ou $cont > \gamma$;
-

De modo semelhante ao que ocorre no *K-means*, o agrupamento final gerado pelo Fuzzy C-Means também é dependente dos parâmetros de inicialização. Desta forma, não há garantia de que, ao obter um agrupamento, este seja o melhor agrupamento possível e o mínimo global tenha sido atingido. Uma alternativa para buscar a otimização dos resultados é executar diferentes inicializações da matriz U_h .

3.2.3 Vizinho mais próximo

Dentro do contexto do agrupamento de ST, em (ZHANG, et al., 2011) uma extensão dos conceitos básicos do algoritmo do vizinho mais próximo é combinada à teoria de grafos, à métrica de similaridade triangular e à distância DTW, com o objetivo de realizar o agrupamento séries temporais.

De acordo com a abordagem apresentada em (ZHANG, et al., 2011), considere um grafo direcionado ponderado $G_w = (N, L, W)$, um conjunto $N = \{n_1, n_2, n_3, \dots, n_N\}$ de nós, um conjunto $L = \{l_1, l_2, l_3, \dots, l_M\}$ de arestas e um conjunto $W = \{w_1, w_2, w_3, \dots, w_M\}$ de pesos para ponderar as M arestas. Cada nó representa uma ST. As arestas, responsáveis por conectar um par de nós n_i e n_j , são denotadas por l_{ij} , tal que $l_{ij} \neq l_{ji}$. A intensidade S_i do i – ésimio nó, é definida pela somatória dos pesos das arestas as quais ele está conectado, conforme apresentado na Equação 3.15.

$$S_i = \sum_{j \in N} w_{ij} \quad (3.15)$$

O peso w_{ij} da aresta l_{ij} representa a similaridade entre os nós i e j , calculado pela similaridade triangular. Assim, a rede de vizinhos mais próximos é construída considerando os k vizinhos com maior semelhança. Considere as semelhanças entre seis ST apresentadas na Tabela 3.1.

Tabela 3.1 – Similaridade entre séries temporais

	n_1	n_2	n_3	n_4	n_5	n_6
n_1	-	0,6	0,2	0,5	0,4	0,3
n_2	0,6	-	0,7	0,1	0,8	0,2
n_3	0,2	0,7	-	0,9	0,8	0,3
n_4	0,8	0,1	0,9	-	0,4	0,2
n_5	0,4	0,6	0,8	0,4	-	0,7
n_6	0,3	0,2	0,3	0,2	0,7	-

No exemplo da Tabela 3.1, ao serem considerados os 2 vizinhos mais próximos, a ilustração do grafo resultante é apresentada na Figura 3.10.

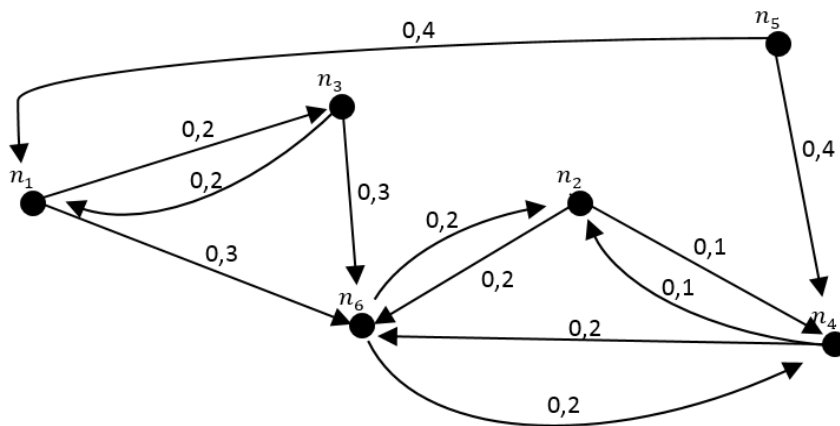


Figura 3.10 – Rede dos vizinhos mais próximos para $k = 2$

Observe que, ao afirmar que um determinado ponto está mais próximo do outro, não podemos deduzir que a recíproca seja verdadeira. Por exemplo. Considere o nó n_1 . Os seus dois vizinhos mais próximos são os nós n_3 e n_6 , respectivamente. Entretanto, o nó n_1 não é um dos dois vizinhos mais próximos de n_6 .

Para gerar o agrupamento dos dados proposto por meio do algoritmo dos k vizinhos mais próximos, deve ser considerado o número de arestas k_i^{in} de cada nó, onde k_i^{in} é a quantidade de arestas que chegam ao i – ésimo nó. Desta forma, de acordo com as premissas apresentadas, o algoritmo de agrupamento dos dados é definido na sequência.

Algoritmo 3.3: k vizinhos mais próximos

1. Enquanto $N \neq \emptyset$
 2. Ordenar os nós do conjunto N , em ordem decrescente considerando o número de arestas k_i^{in} ;
 3. Escolha o nó n com maior quantidade de arestas k_i^{in} ;
 4. Adicione o nó n ao conjunto C e remova-o de N ;
 5. Identifique os vizinhos de n , remova-os de N e adicione-os ao conjunto EC ;
 6. Considerando os nós restantes em N , retornar ao Passo 1;
 7. Defina C como os centros de *cluster*.
-

No Algoritmo 3.3, o conjunto C representa os centros de *clusters* e EC é o conjunto que possui os respectivos elementos de cada um dos *clusters*. Em bases de dados com uma grande quantidade de amostras, este método reduz consideravelmente o custo computacional para definição do agrupamento. Predominantemente, o uso do algoritmo k -vizinhos mais próximos para agrupamento de ST está diretamente associado ao uso da distância DTW e suas extensões (GÓRECKI, ŁUCZAK, 2014) (KURBALIJA, et al., 2014) (YU, et al., 2011) e (TORMENE, et al., 2009).

3.2.4 Classificação comportamental das séries temporais

Com o enfoque do agrupamento dinâmico das séries temporais, D'Urso e Maharaj (2009) utilizaram o algoritmo Fuzzy C-Means para agrupar as séries por meio do coeficiente de correlação dos dados observados nas ST. Em (JI, et al., 2013) os autores apresentam uma extensão ao Fuzzy C-Means com a introdução do que é denominado “*key point*”, utilizado para realizar o agrupamento dinâmico. Em ambas abordagens os autores consideram que a mudança comportamental em uma ST pode ocorrer de três formas. A primeira delas está relacionada a mudança nos valores da série, representada na Figura 3.11.

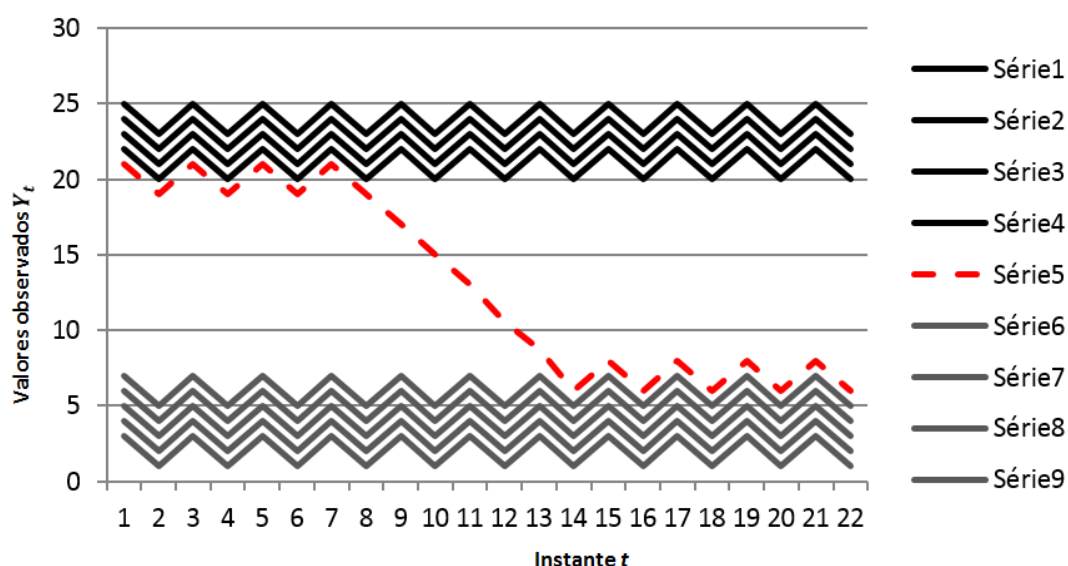


Figura 3.11 – Mudança nos valores observados na série temporal.
Adaptado de (D'URSO; MAHARAJ, 2009)

A Figura 3.11 apresenta uma mudança em uma série temporal onde o comportamento da série permanece semelhante ao observado no período anterior. Assim, a alteração está relacionada apenas aos valores reais observados na Série 5.

Na Figura 3.12 é apresentado um exemplo onde a alteração está relacionada ao comportamento da série, pois os valores observados após a alteração ainda continuam dentro do intervalo observado antes da alteração.

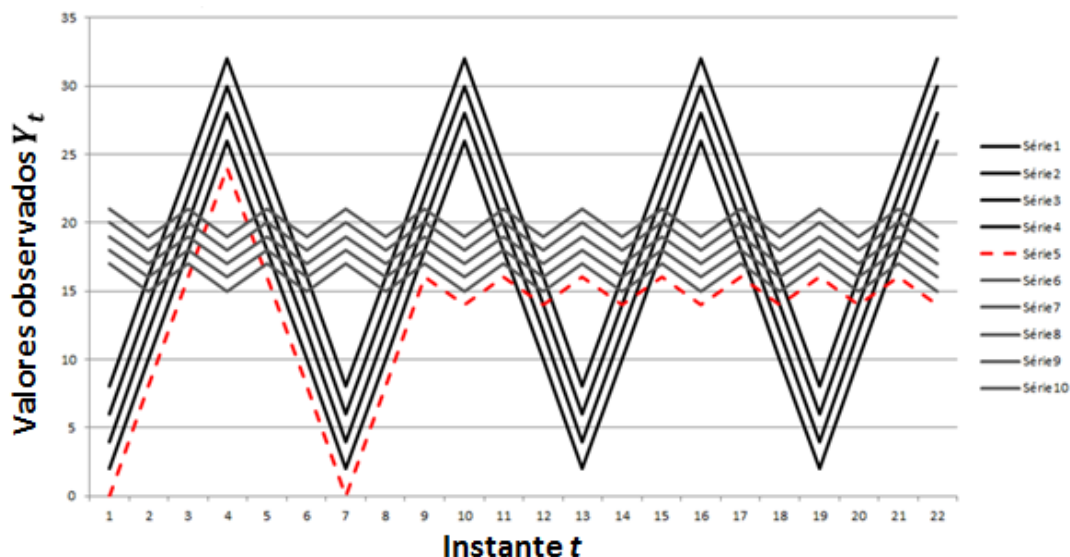


Figura 3.12 – Mudança no formato da série temporal.
Adaptado de (D'URSO; MAHARAJ, 2009)

O terceiro tipo de mudança possível em uma ST, é quando as alterações ocorrem tanto no formato, quanto nos valores observados na série, não apresentando relação alguma com o seu comportamento anterior. A Figura 3.13 ilustra um exemplo deste tipo de alteração.

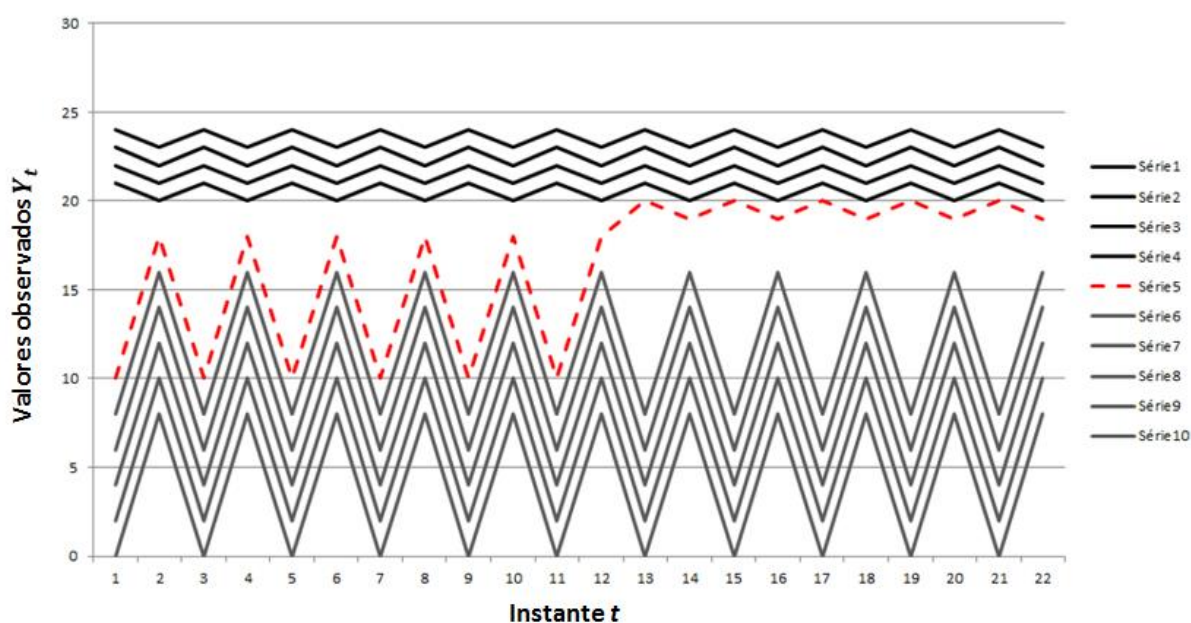


Figura 3.13 – Mudança nos valores e no formato da série temporal.
Adaptado de (D'URSO; MAHARAJ, 2009)

A análise das séries temporais proposta neste trabalho tem como parâmetro principal os valores das amostras ao longo do período de observação que, ao ser combinada com a segmentação da série, é capaz de identificar mudanças no padrão comportamental das ST com base nos valores observados.

3.3 Considerações finais

Neste capítulo foram apresentadas as principais métricas utilizadas para o agrupamento de ST, em especial o algoritmo DTW, amplamente utilizado na literatura. Dentre os algoritmos de agrupamento apresentados, o algoritmo do vizinho mais próximo tem se mostrado eficaz na classificação de ST quando associado ao algoritmo DTW. Entretanto, recentemente o FCM tem ganhado espaço na análise de ST para o agrupamento dinâmico.

No próximo capítulo serão apresentadas as abordagens desenvolvidas para o pré-processamento das ST e para a previsão de valores. As abordagens desenvolvidas apresentam uma extensão das RLF para a representação do conhecimento nas STF, uma nova técnica de alisamento e a combinação de conceitos clássicos das ST com os conceitos das séries temporais fuzzy para o pré-processamento e para a previsão de valores.

Capítulo 4

ANÁLISE DE SÉRIES TEMPORAIS FUZZY PARA PREVISÃO DE VALORES

Durante o período de pesquisa para elaboração desta tese, diferentes oportunidades foram identificadas para o desenvolvimento de abordagens de previsão de valores com o uso de STF (SONG; CHISSOM, 1993a), (SONG; CHISSOM, 1993b) (SONG; CHISSOM, 1994). Por tratar-se de uma área pouco explorada, ainda há muito a ser desenvolvido neste campo e, por esta razão, diversos modelos de previsão foram testados ao longo da pesquisa.

A primeira necessidade identificada durante a pesquisa foi a de utilizar um modelo de pré-processamento antes de realizar qualquer a previsão. Foram realizados diversos experimentos utilizando modelos fuzzy de previsão que, na proposta original, não aplicam nenhum tipo de pré-processamento nos dados. Os resultados obtidos comprovaram que, ao ser realizado o pré-processamento nos dados, o modelo de previsão apresenta uma melhora significativa em sua acurácia. Os modelos de previsão desenvolvidos nesta tese são apresentados na Seção 4.1.

A partir dos estudos com foco nos métodos clássicos de previsão, foi identificada uma carência de trabalhos que combinem os conceitos clássicos de séries tem-

porais com os conceitos de séries temporais fuzzy. Assim, a partir de uma nova técnica de alisamento inspirada no AES, a principal abordagem de previsão desenvolvida tem como objetivo apresentar um modelo fuzzy para previsão capaz de atribuir maior influência ao comportamento mais recente da série, mas que também seja flexível ao ponto de permitir que, quando necessário, seja atribuída maior influência ao comportamento mais antigo da ST. Para melhorar a acurácia das previsões, foram introduzidos os conceitos de regressão linear simples aplicados às séries temporais fuzzificadas (SANTOS; CAMARGO, NO PRELO).

Antes de ser definida a abordagem principal para previsões, foram realizados diversos experimentos que resultaram em outras duas abordagens para a previsão de valores. A primeira delas, descrita na Seção 4.2.1, define grupos de RLF a partir de um intervalo temporal calculado de acordo com a quantidade de dados observados na ST. Para cada grupo de RLF é atribuído um grau de influência no cálculo do valor a ser previsto (SANTOS; CAMARGO, 2014b). A outra abordagem desenvolvida, apresentada na Seção 4.2.2, apresenta uma extensão das RLF combinada com os conceitos do alisamento exponencial simples (SANTOS; CAMARGO, 2014a).

Com o objetivo apresentar a evolução da pesquisa, neste capítulo serão descritos, além do de pré-processamento desenvolvido, os três modelos de previsão resultantes da pesquisa. Os resultados dos experimentos realizados são apresentados no Capítulo 6.

4.1 Pré-processamento das séries temporais fuzzy

Nesta seção serão apresentados os três modelos de pré-processamento desenvolvidos e utilizados tanto no processo de previsão, quanto na análise comportamental das séries. O primeiro modelo foi desenvolvido com o objetivo de verificar se o uso do pré-processamento realmente implica em uma melhora da acurácia nos processos de previsão. As duas outras abordagens de pré-processamento foram desenvolvidas a partir de ajustes realizados na primeira abordagem.

4.1.1 Pré-processamento: primeira abordagem

Embora diversos trabalhos com STF não realizam o pré-processamento dos dados (CHENG, et al., 2008) (JOSHI; KUMAR, 2013), (KOROL, 2014) e (QIU, et al., 2011) esse processo é de fundamental importância para a melhora na acurácia das previsões e para um agrupamento mais adequado das ST. As definições do domínio, o número de termos linguísticos e as funções de pertinência dos conjuntos fuzzy, possuem influência direta na acurácia do modelo de previsão.

O modelo de pré-processamento aqui descrito (SANTOS; CAMARGO, 2013), tem como principal objetivo analisar os dados para permitir uma melhor representação fuzzy dos valores *crisp* observados em uma ST. Para isso, a abordagem desenvolvida combina alguns procedimentos disponíveis na literatura para a definição de um novo modelo. O processo desenvolvido está dividido em quatro etapas, apresentadas na Figura 4.1.

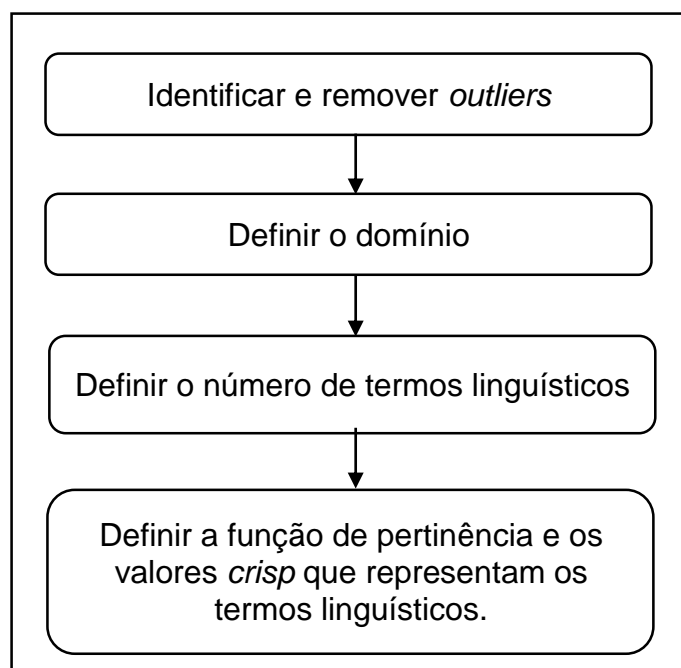


Figura 4.1 – Pré-processamento das ST

A existência de *outliers* na base de dados pode exercer uma influência negativa em todo o restante do processo de análise das séries. Assim, o primeiro passo é a remoção dos *outliers*. A técnica utilizada para identificar e remover os *outliers* é a mesma descrita na Seção 2.6 (BARNETT; LEWIS, 1994).

Na sequência, o método deve definir um universo de discurso apropriado para a ST. Assim, após a identificação e remoção dos *outliers*, o universo de discurso da ST é definido pela Equação 4.1, onde y_{min} e y_{max} são, respectivamente, o menor e o maior valores da ST e σ é o desvio padrão dos dados observados.

$$D = [y_{min} - \sigma, y_{max} + \sigma] \quad (4.1)$$

Em relação a quantidade de termos linguísticos a serem utilizados na representação das séries temporais, a abordagem de pré-processamento proposta nesta seção utiliza a técnica descrita entre os passos 1 e 6, apresentados na Seção 2.6 (TANUWIJAYA; CHEN, 2009b).

Conjuntos fuzzy com tamanhos diferentes em uma mesma variável, podem representar melhor a estrutura interna de dados desigualmente distribuídos sobre o universo de discurso e, portanto, produzir melhores resultados de previsão (SANTOS; CAMARGO, 2013). Sendo assim, a quantidade c de termos linguísticos definido pelos 6 passos descritos na Seção 2.6 (TANUWIJAYA; CHEN, 2009b) é utilizada como parâmetro de entrada para o algoritmo de agrupamento Fuzzy C-Means, descrito na Seção 3.2.2 (BEZDEK, et al., 1992).

Após a execução do Fuzzy C-Means, os centroides encontrados são utilizados na representação dos termos linguísticos da ST, ou seja, o valor de representação numérica do termo linguístico L_k corresponde ao valor do centróide k obtido por meio do Fuzzy C-Means. Assim, a definição dos parâmetros dos conjuntos fuzzy é realizada como segue:

- Os termos linguísticos L_k , onde $1 < k < c$, são representados por conjuntos fuzzy triangulares;
- Os parâmetros esquerdo e direito que definem os limites dos conjuntos fuzzy triangulares L_k , são definidos, respectivamente, pelos valores de $centro_cluster_{k-1}$ e de $centro_cluster_{k+1}$ fornecidos pelo Fuzzy C-Means;
- Para o termo linguístico L_1 a representação é por meio do conjunto fuzzy trapezoidal com $pertinência = 1$ no intervalo $[y_{min} - \sigma, centro_cluster_1]$, onde σ é o desvio padrão dos dados observados na ST;

- Os parâmetros esquerdo e direito de L_1 são definidos, respectivamente, por $y_{min} - \sigma$ e $centro_cluster_2$;
- Para o termo linguístico L_c , a representação também será por meio do conjunto fuzzy trapezoidal, com $pertinência = 1$ no intervalo $[centro_cluster_c, y_{max} + \sigma,]$, onde σ é o desvio padrão da série;
- Os parâmetros esquerdo e direito de L_c são definidos, respectivamente, por $centro_cluster_{c-1}$ e por $y_{max} + \sigma$;

Após a definição do suporte de cada um dos conjuntos fuzzy e, conseqüentemente, dos valores de representação dos termos linguísticos, devem ser realizadas a fuzzificação da ST e a definição da base de conhecimento do sistema por meio da identificação das RLF.

4.1.2 Pré-processamento: segunda abordagem

O pré-processamento descrito nesta seção apresenta apenas uma alteração em relação ao pré-processamento apresentado na Seção 4.1.1. Esta alteração está relacionada a definição do *threshold* τ da Equação 2.15, que é calculado a partir da diferença entre os valores observados na série ordenados de forma crescente.

Em ST longas e com um desvio padrão relativamente alto, o *threshold* definido pela Equação 2.15 implica em um número elevado de termos linguísticos para representar a ST. Em séries com 100 amostras, por exemplo, alguns experimentos chegaram a indicar 60 termos linguísticos para a representação fuzzy da série.

Como o parâmetro τ possui influência direta na definição do número de termos, a alternativa encontrada foi aumentar o valor deste parâmetro somando o desvio padrão da diferença das amostras. Assim, novo valor do *threshold* é calculado a partir da Equação 4.2 aplicada aos dados da série temporal ordenados de forma crescente.

$$\tau = \frac{\sum_{i=1}^{n-1} (y_{i+1} - y_i)}{n - 1} + \sigma \quad (4.2)$$

onde y_i é o i - ésimo termos da ST ordenada de forma crescente, σ é o desvio padrão da diferença das amostras da ST ordenada em uma sequência crescente. O restante

do processo para definir o número de termos linguísticos é o mesmo utilizado no pré-processamento apresentado na Seção 4.1.1.

4.1.3 Pré-processamento: terceira abordagem

Nas abordagens descritas anteriormente, considerando que foram definidos c termos linguísticos, após a execução do Fuzzy C-Means os conjuntos fuzzy são definidos como ilustrado na Figura 4.2, onde σ é o desvio padrão das amostras disponíveis e, y_{min} e y_{max} são, respectivamente, o menor e o maior valores *crisp* observados nas amostras disponíveis da ST.

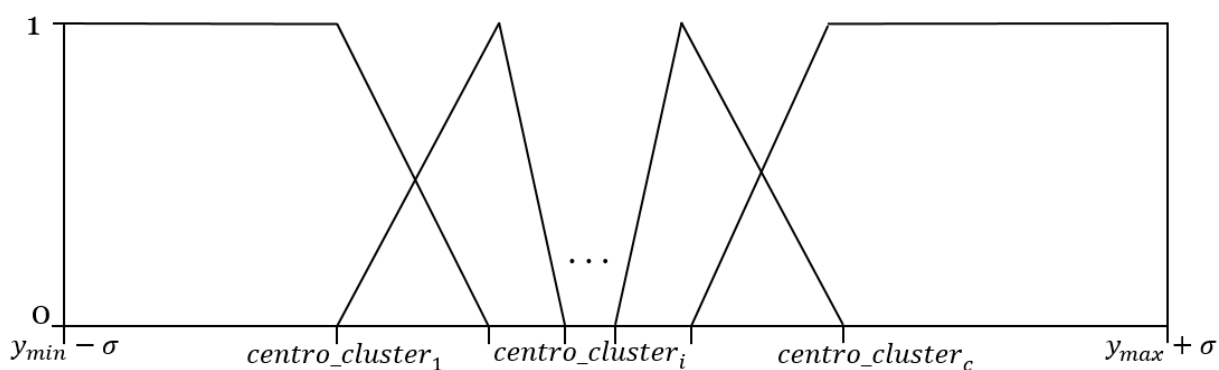


Figura 4.2 – Definição do domínio e dos conjuntos fuzzy conforme pré-processamento apresentado na Seção 4.1.2

A razão pela qual o desvio padrão é considerado na definição do domínio é, justamente, para permitir que valores não observados até o momento do pré-processamento, possam ser representados pelo modelo em uma previsão. Entretanto, como é possível observar na Figura 4.2, não há uma representação linguística adequada para os valores que possam surgir acima do $centro_cluster_c$, ou abaixo do $centro_cluster_1$, o que acaba prejudicando a acurácia em problemas onde há previsões de valores nestes intervalos.

Para sanar este problema, a abordagem de pré-processamento apresentada nesta seção realiza a adição de novos conjuntos fuzzy nos intervalos $[y_{min} - \sigma, centro_cluster_1]$ e $[centro_cluster_c, y_{max} + \sigma]$, conforme apresentado a seguir:

- **Passo 1.** Calcular a média aritmética da diferença $media_{dif}$ entre $centro_cluster_i$ e $centro_cluster_{i+1}$, para $i = 1, 2, \dots, c - 1$.

- **Passo 2.** Calcular o número de conjuntos adicionais ca_{inf} no limite inferior $[y_{min} - \sigma, centro_cluster_1]$, como apresentado na Equação 4.3.

$$ca_{inf} = \left\lfloor \frac{centro_cluster_1 - (y_{min} - \sigma)}{media_{dif}} \right\rfloor \quad (4.3)$$

- **Passo 3.** Definir o parâmetro central de cada novo conjunto fuzzy triangular adicionado no intervalo $[y_{min} - \sigma, centro_cluster_1]$, a partir da expressão $centro_cluster_1 - media_{dif} * add$, onde $add = 1, 2, \dots, ca_{inf}$.
- **Passo 4.** Se $1 < add < ca_{inf}$, definir os parâmetros esquerdo e direito do novo conjunto triangular, respectivamente, por $centro_cluster_1 - media_{dif} * (add + 1)$ e $centro_cluster_1 - media_{dif} * (add - 1)$.
- **Passo 5.** Se $add = 1$, definir os parâmetros esquerdo e direito do novo conjunto triangular, respectivamente, por $centro_cluster_1 - media_{dif} * 2$ e $centro_cluster_1$.
- **Passo 6.** Se $add = ca_{inf}$, definir os parâmetros esquerdo e direito do novo conjunto triangular, respectivamente, por $y_{min} - \sigma$ e $centro_cluster_1 - media_{dif} * (add - 1)$.
- **Passo 7.** Para os conjuntos fuzzy triangulares adicionados no intervalo $[centro_cluster_c, y_{max} + \sigma]$, calcular o número de conjuntos adicionais ca_{sup} como apresentado na Equação 4.4.

$$ca_{sup} = \left\lfloor \frac{(y_{max} + \sigma) - centro_cluster_c}{media_{dif}} \right\rfloor \quad (4.4)$$

- **Passo 8.** De modo análogo ao descrito para o intervalo $[y_{min} - \sigma, centro_cluster_1]$ nos passos de 3 à 6, definir os conjuntos fuzzy adicionais a partir das variáveis ca_{sup} , $media_{dif}$ e $centro_cluster_c$. Para definir os novos conjuntos adicione $media_{dif}$ ao valor $centro_cluster_c$ na definição do parâmetro central de cada novo conjunto triangular até o limite superior do domínio, definido por $y_{max} + \sigma$.

Em problemas como o ilustrado na Figura 4.2, a adição dos novos conjuntos triangulares fuzzy como descrito entre os passos 1 e 8 melhora significativamente a acurácia dos valores previstos. Sem a alteração no pré-processamento realizada nesta abordagem, todos os valores previstos no cenário apresentado pela Figura 4.3 seriam representados pelo valor *crisp* do $centro_cluster_c$ ou pelo limite superior do domínio, definido por $y_{max} + \sigma$, o que certamente prejudicaria a acurácia das previsões.

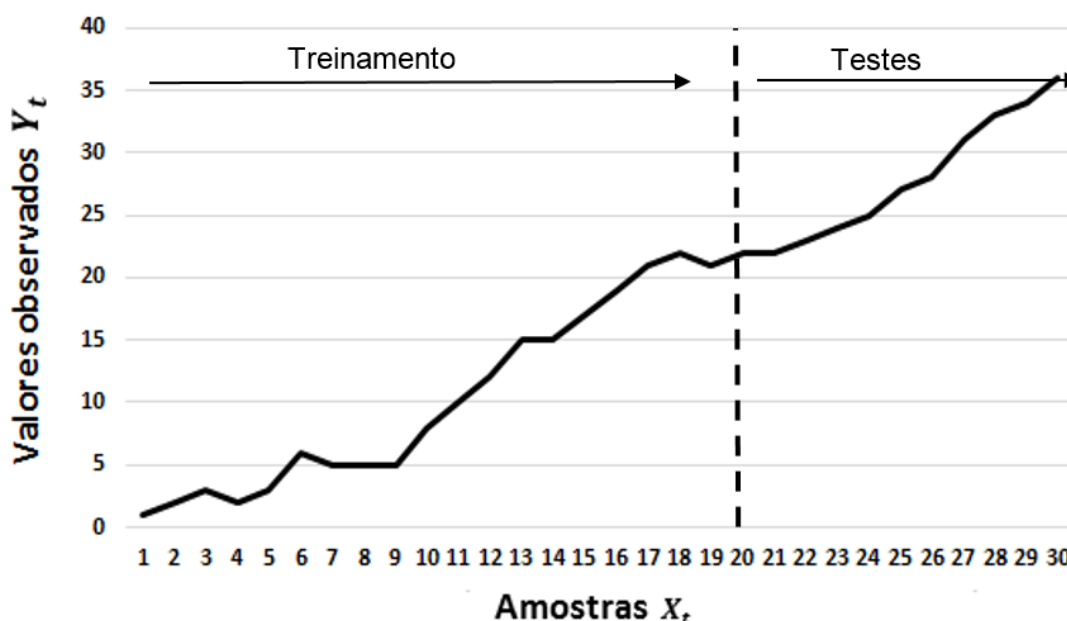


Figura 4.3 – Exemplo de série temporal

Outra característica importante em problemas como o da Figura 4.3 é a atualização constante da base de conhecimento. Após a chegada de novas amostras, novas RLF devem ser adicionadas à base, o que permite ao modelo continuar aprendendo mesmo depois do treinamento inicial.

4.2 Modelos de previsão

Nesta seção serão apresentadas as três principais abordagens de previsão desenvolvidas durante a pesquisa.

4.2.1 Previsões em STF baseadas no intervalo temporal das amostras

A primeira abordagem de previsão utiliza como técnica de pré-processamento o modelo apresentado na Seção 4.1.2. A abordagem desenvolvida nesta seção tem como principal característica classificar o conhecimento extraído da ST em diferentes grupos temporais (SANTOS; CAMARGO, 2014b). O objetivo é dar maior influência no cálculo do valor previsto ao conhecimento pertencente ao grupo temporal com o conhecimento mais recente da série. Para isso, é realizada a classificação temporal das relações lógicas fuzzy.

A base de conhecimento do sistema é composta por RLF de segunda ordem, ou seja, com dois antecedentes na relação de implicação, como apresentado na Expressão 4.5.

$$L_i, L_{i+1} \rightarrow L_{i+2}, \quad i = 1, 2, 3, \dots, n - 2 \quad (4.5)$$

onde n é o número de amostras da ST e L_i é o termo linguístico que representa o valor observado no instante i . Considerando as RLF de segunda ordem como apresentado na Expressão 4.5, ou seja, com dois antecedentes, as duas últimas amostras da ST indicarão quais RLF serão utilizadas no cálculo do valor a ser previsto. Assim, serão consideradas todas as implicações das RLF na base de conhecimento que tiverem os antecedentes $L_i = L_{n-1}$ e $L_{i+1} = L_n$.

A escolha por RLF de segunda ordem deve-se ao fato de que o uso das relações de primeira ordem para a representação do conhecimento implica em uma quantidade excessiva de RLF no cálculo do valor previsto, o que acaba ocasionando um desvio significativo nas previsões quando comparadas com os valores reais. Esse desvio também ocorre ao utilizarmos RLF de terceira ordem ou superior pois, ao contrário do que ocorre quando são utilizadas RLF de primeira ordem, o número de termos linguísticos utilizados no cálculo é insuficiente para a identificação da tendência da série temporal.

A abordagem apresentada nesta seção considera a existência de três grupos temporais de RLF (G_1 , G_2 e G_3), definidos a partir das n amostras da série conforme descrito na sequência:

- **Passo 1.** A partir da ST fuzzificada, construa uma base de RLF de 2ª ordem como apresentado na Expressão 4.5, com um total de relações r igual a $n - 2$.
- **Passo 2.** Considere as relações r_j , onde $j = 1, 2, \dots, n - 2$, definidas no passo anterior. Se $j \leq (n - 2)/3$, então $r_j \in G_1$. Se $(n - 2)/3 < j \leq (n - 2)/3 \times 2$, então $r_j \in G_2$. Se $j > (n - 2)/3 \times 2$, então $r_j \in G_3$.

O grau de influência, ou o peso p_{G_i} , do grupo de RLF G_i , onde, $i = 1, 2$ e 3 , é definido empiricamente respeitando a restrição $p_{G_1} < p_{G_2} < p_{G_3}$. Assim, o valor previsto é calculado pela Equação 4.6.

$$V_p = \sum_{i=1}^3 V_{G_i} \times p_{G_i} \quad (4.6)$$

onde V_p é o valor previsto e V_{G_i} é o valor calculado a partir das RLF presentes no Grupo i , conforme Equação 4.7, que considera o número de ocorrências de cada RLF dentro do grupo.

$$V_{G_i} = \frac{\sum_{j=1}^l Def(L_j) \times f_{L_j}}{\sum_{j=1}^l f_{L_j}} \quad (4.7)$$

onde $i = 1, 2$ e 3 , l indica o número de diferentes RLF presentes no Grupo i , f_{L_j} é o número de ocorrências de RLFs com L_j no consequente dentro do Grupo i , ou seja, quantas vezes a relação $L_{n-1}, L_n \rightarrow L_j$ ocorre dentro do grupo e, $Def(L_j)$ é o parâmetro central do conjunto triangular que representa o termo linguístico L_j .

Para ilustrar como o cálculo do valor previsto é realizado nesta abordagem de previsão, considere a série temporal com os índices *Taiwan Stock Exchange Capitalization Weighted Stock Index* (TAIEX) da bolsa de valores de Taiwan para o ano de 2001, utilizada nos experimentos apresentados no Capítulo 6. Os índices entre janeiro e outubro de 2001 foram utilizados para treinamento do modelo e os testes foram realizados utilizando os índices de novembro e dezembro do mesmo ano. Após o pré-processamento nos dados de treinamento, os valores *crisp* obtidos para cada *cluster* e, conseqüentemente, para a representação dos termos linguísticos na série temporal fuzzificada, são apresentados na Tabela 4.1.

Tabela 4.1 – Valores *crisp* de representação para cada termo linguístico do índice TAIEX 2001

<i>Cluster</i>	Valor	<i>Cluster</i>	Valor	<i>Cluster</i>	Valor	<i>Cluster</i>	Valor
0	3488,58	13	3915,35	26	4410,30	39	4837,79
1	3525,41	14	3952,83	27	4453,63	40	4895,68
2	3567,26	15	3987,10	28	4473,64	41	4979,42
3	3588,63	16	4012,67	29	4489,39	42	5033,78
4	3622,13	17	4042,43	30	4509,75	43	5084,22
5	3637,01	18	4065,64	31	4528,65	44	5150,37
6	3706,00	19	4145,09	32	4552,98	45	5175,34
7	3778,23	20	4184,87	33	4591,49	46	5227,90
8	3791,01	21	4220,44	34	4629,14	47	5309,69
9	3800,05	22	4301,64	35	4659,84	48	5406,05
10	3814,66	23	4335,51	36	4706,31	49	5481,84
11	3848,60	24	4354,66	37	4741,69	50	5588,32
12	3900,91	25	4373,35	38	4780,37	51	5801,55

Após a fuzzificação da ST, 199 relações lógicas fuzzy, como a apresentada na Expressão 4.5, foram identificadas e divididas em 3 grupos (G_1 , G_2 e G_3) que compõem a base de conhecimento, sendo os Grupos 1 e 2 com 66 RLF cada um e o Grupo 3 com 67 relações. As mais antigas, pertencentes ao *Grupo 1*, são apresentadas na Tabela 4.2. As RLF do *Grupo 2*, são apresentadas na Tabela 4.3, e as RLF do *Grupo 3*, que representam o comportamento mais recente da ST, são apresentadas na Tabela 4.4.

Tabela 4.2 – RLF de 2ª ordem pertencentes ao *Grupo 1* da série temporal TAIEX 2001

RLF do <i>Grupo 1</i>				
$L_{40}, L_{40} \rightarrow L_{44}$	$L_{51}, L_{51} \rightarrow L_{51}$	$L_{51}, L_{51} \rightarrow L_{51}$	$L_{50}, L_{49} \rightarrow L_{49}$	$L_{51}, L_{51} \rightarrow L_{50}$
$L_{40}, L_{44} \rightarrow L_{47}$	$L_{51}, L_{51} \rightarrow L_{51}$	$L_{51}, L_{51} \rightarrow L_{51}$	$L_{50}, L_{50} \rightarrow L_{50}$	$L_{51}, L_{50} \rightarrow L_{49}$
$L_{44}, L_{47} \rightarrow L_{45}$	$L_{51}, L_{51} \rightarrow L_{51}$	$L_{51}, L_{51} \rightarrow L_{51}$	$L_{50}, L_{50} \rightarrow L_{51}$	$L_{50}, L_{49} \rightarrow L_{48}$
$L_{47}, L_{45} \rightarrow L_{48}$	$L_{51}, L_{51} \rightarrow L_{51}$	$L_{51}, L_{51} \rightarrow L_{51}$	$L_{50}, L_{51} \rightarrow L_{51}$	$L_{49}, L_{48} \rightarrow L_{49}$
$L_{45}, L_{48} \rightarrow L_{48}$	$L_{51}, L_{51} \rightarrow L_{50}$	$L_{51}, L_{51} \rightarrow L_{50}$	$L_{51}, L_{51} \rightarrow L_{50}$	$L_{48}, L_{49} \rightarrow L_{48}$
$L_{48}, L_{48} \rightarrow L_{48}$	$L_{51}, L_{50} \rightarrow L_{51}$	$L_{51}, L_{50} \rightarrow L_{49}$	$L_{51}, L_{50} \rightarrow L_{50}$	$L_{49}, L_{48} \rightarrow L_{47}$
$L_{48}, L_{48} \rightarrow L_{47}$	$L_{50}, L_{51} \rightarrow L_{51}$	$L_{50}, L_{49} \rightarrow L_{49}$	$L_{50}, L_{50} \rightarrow L_{50}$	$L_{48}, L_{47} \rightarrow L_{49}$
$L_{48}, L_{47} \rightarrow L_{48}$	$L_{51}, L_{51} \rightarrow L_{51}$	$L_{49}, L_{49} \rightarrow L_{50}$	$L_{50}, L_{50} \rightarrow L_{51}$	$L_{47}, L_{49} \rightarrow L_{49}$
$L_{47}, L_{48} \rightarrow L_{50}$	$L_{51}, L_{51} \rightarrow L_{51}$	$L_{49}, L_{50} \rightarrow L_{50}$	$L_{50}, L_{51} \rightarrow L_{51}$	$L_{49}, L_{49} \rightarrow L_{49}$
$L_{48}, L_{50} \rightarrow L_{51}$	$L_{51}, L_{51} \rightarrow L_{51}$	$L_{50}, L_{50} \rightarrow L_{51}$	$L_{51}, L_{51} \rightarrow L_{51}$	$L_{49}, L_{49} \rightarrow L_{48}$
$L_{50}, L_{51} \rightarrow L_{51}$	$L_{51}, L_{51} \rightarrow L_{51}$	$L_{50}, L_{51} \rightarrow L_{51}$	$L_{51}, L_{51} \rightarrow L_{51}$	-
$L_{51}, L_{51} \rightarrow L_{50}$	$L_{51}, L_{51} \rightarrow L_{51}$	$L_{51}, L_{51} \rightarrow L_{50}$	$L_{51}, L_{51} \rightarrow L_{51}$	-
$L_{51}, L_{50} \rightarrow L_{51}$	$L_{51}, L_{51} \rightarrow L_{51}$	$L_{51}, L_{50} \rightarrow L_{50}$	$L_{51}, L_{51} \rightarrow L_{51}$	-
$L_{50}, L_{51} \rightarrow L_{51}$	$L_{51}, L_{51} \rightarrow L_{51}$	$L_{50}, L_{50} \rightarrow L_{50}$	$L_{51}, L_{51} \rightarrow L_{51}$	-

Tabela 4.3 – RLF de 2ª ordem pertencentes ao Grupo 2 da série temporal TAIEX 2001

RLF do Grupo 2				
$L_{48}, L_{48} \rightarrow L_{49}$	$L_{46}, L_{45} \rightarrow L_{46}$	$L_{45}, L_{43} \rightarrow L_{43}$	$L_{46}, L_{44} \rightarrow L_{44}$	$L_{37}, L_{36} \rightarrow L_{36}$
$L_{48}, L_{49} \rightarrow L_{50}$	$L_{45}, L_{46} \rightarrow L_{44}$	$L_{43}, L_{43} \rightarrow L_{42}$	$L_{44}, L_{44} \rightarrow L_{43}$	$L_{36}, L_{36} \rightarrow L_{35}$
$L_{49}, L_{50} \rightarrow L_{50}$	$L_{46}, L_{44} \rightarrow L_{46}$	$L_{43}, L_{42} \rightarrow L_{42}$	$L_{44}, L_{43} \rightarrow L_{42}$	$L_{36}, L_{35} \rightarrow L_{34}$
$L_{50}, L_{50} \rightarrow L_{50}$	$L_{44}, L_{46} \rightarrow L_{45}$	$L_{42}, L_{42} \rightarrow L_{42}$	$L_{43}, L_{42} \rightarrow L_{42}$	$L_{35}, L_{34} \rightarrow L_{32}$
$L_{50}, L_{50} \rightarrow L_{50}$	$L_{46}, L_{45} \rightarrow L_{44}$	$L_{42}, L_{42} \rightarrow L_{41}$	$L_{42}, L_{42} \rightarrow L_{41}$	$L_{34}, L_{32} \rightarrow L_{34}$
$L_{50}, L_{50} \rightarrow L_{49}$	$L_{45}, L_{44} \rightarrow L_{43}$	$L_{42}, L_{41} \rightarrow L_{43}$	$L_{42}, L_{41} \rightarrow L_{40}$	$L_{32}, L_{34} \rightarrow L_{29}$
$L_{50}, L_{49} \rightarrow L_{49}$	$L_{44}, L_{43} \rightarrow L_{44}$	$L_{50}, L_{49} \rightarrow L_{49}$	$L_{41}, L_{40} \rightarrow L_{38}$	$L_{34}, L_{29} \rightarrow L_{25}$
$L_{49}, L_{49} \rightarrow L_{48}$	$L_{43}, L_{44} \rightarrow L_{43}$	$L_{41}, L_{43} \rightarrow L_{46}$	$L_{40}, L_{38} \rightarrow L_{39}$	$L_{29}, L_{25} \rightarrow L_{25}$
$L_{49}, L_{48} \rightarrow L_{48}$	$L_{44}, L_{43} \rightarrow L_{41}$	$L_{43}, L_{46} \rightarrow L_{44}$	$L_{38}, L_{39} \rightarrow L_{38}$	$L_{25}, L_{25} \rightarrow L_{21}$
$L_{48}, L_{48} \rightarrow L_{47}$	$L_{43}, L_{41} \rightarrow L_{41}$	$L_{46}, L_{44} \rightarrow L_{46}$	$L_{39}, L_{38} \rightarrow L_{40}$	$L_{49}, L_{48} \rightarrow L_{48}$
$L_{48}, L_{47} \rightarrow L_{48}$	$L_{41}, L_{41} \rightarrow L_{46}$	$L_{44}, L_{46} \rightarrow L_{47}$	$L_{38}, L_{40} \rightarrow L_{40}$	-
$L_{47}, L_{48} \rightarrow L_{46}$	$L_{41}, L_{46} \rightarrow L_{46}$	$L_{46}, L_{47} \rightarrow L_{46}$	$L_{40}, L_{40} \rightarrow L_{39}$	-
$L_{48}, L_{46} \rightarrow L_{46}$	$L_{46}, L_{46} \rightarrow L_{45}$	$L_{47}, L_{46} \rightarrow L_{46}$	$L_{40}, L_{39} \rightarrow L_{37}$	-
$L_{46}, L_{46} \rightarrow L_{45}$	$L_{46}, L_{45} \rightarrow L_{43}$	$L_{46}, L_{46} \rightarrow L_{44}$	$L_{39}, L_{37} \rightarrow L_{36}$	-

Tabela 4.4 – RLF de 2ª ordem pertencentes ao Grupo 3 da série temporal TAIEX 2001

RLF do Grupo 3				
$L_{20}, L_{21} \rightarrow L_{19}$	$L_{27}, L_{28} \rightarrow L_{31}$	$L_{30}, L_{30} \rightarrow L_{27}$	$L_3, L_1 \rightarrow L_0$	$L_6, L_8 \rightarrow L_{10}$
$L_{21}, L_{19} \rightarrow L_{17}$	$L_{28}, L_{31} \rightarrow L_{33}$	$L_{30}, L_{27} \rightarrow L_{29}$	$L_1, L_0 \rightarrow L_4$	$L_8, L_{10} \rightarrow L_{10}$
$L_{19}, L_{17} \rightarrow L_{19}$	$L_{31}, L_{33} \rightarrow L_{34}$	$L_{27}, L_{29} \rightarrow L_{26}$	$L_0, L_4 \rightarrow L_2$	$L_{10}, L_{10} \rightarrow L_{11}$
$L_{17}, L_{19} \rightarrow L_{22}$	$L_{33}, L_{34} \rightarrow L_{36}$	$L_{29}, L_{26} \rightarrow L_{23}$	$L_4, L_2 \rightarrow L_5$	$L_{10}, L_{11} \rightarrow L_{12}$
$L_{19}, L_{22} \rightarrow L_{23}$	$L_{34}, L_{36} \rightarrow L_{34}$	$L_{26}, L_{23} \rightarrow L_{22}$	$L_2, L_5 \rightarrow L_0$	$L_{11}, L_{12} \rightarrow L_{11}$
$L_{22}, L_{23} \rightarrow L_{24}$	$L_{36}, L_{34} \rightarrow L_{32}$	$L_{23}, L_{22} \rightarrow L_{22}$	$L_5, L_0 \rightarrow L_0$	$L_{12}, L_{11} \rightarrow L_{15}$
$L_{23}, L_{24} \rightarrow L_{24}$	$L_{34}, L_{32} \rightarrow L_{32}$	$L_{22}, L_{22} \rightarrow L_{20}$	$L_0, L_0 \rightarrow L_0$	$L_{11}, L_{15} \rightarrow L_{16}$
$L_{24}, L_{24} \rightarrow L_{29}$	$L_{32}, L_{32} \rightarrow L_{29}$	$L_{22}, L_{20} \rightarrow L_{14}$	$L_0, L_0 \rightarrow L_3$	$L_{15}, L_{16} \rightarrow L_{17}$
$L_{24}, L_{29} \rightarrow L_{31}$	$L_{32}, L_{29} \rightarrow L_{27}$	$L_{20}, L_{14} \rightarrow L_7$	$L_0, L_3 \rightarrow L_1$	$L_{16}, L_{17} \rightarrow L_{18}$
$L_{29}, L_{31} \rightarrow L_{28}$	$L_{29}, L_{27} \rightarrow L_{22}$	$L_{14}, L_7 \rightarrow L_7$	$L_3, L_1 \rightarrow L_4$	$L_{25}, L_{21} \rightarrow L_{20}$
$L_{31}, L_{28} \rightarrow L_{26}$	$L_{27}, L_{22} \rightarrow L_{25}$	$L_7, L_7 \rightarrow L_6$	$L_3, L_1 \rightarrow L_4$	$L_{21}, L_{20} \rightarrow L_{21}$
$L_{28}, L_{26} \rightarrow L_{30}$	$L_{22}, L_{25} \rightarrow L_{25}$	$L_7, L_6 \rightarrow L_3$	$L_4, L_8 \rightarrow L_9$	-
$L_{26}, L_{30} \rightarrow L_{27}$	$L_{25}, L_{25} \rightarrow L_{30}$	$L_{23}, L_{22} \rightarrow L_{22}$	$L_8, L_9 \rightarrow L_6$	-
$L_{30}, L_{27} \rightarrow L_{28}$	$L_{30}, L_{30} \rightarrow L_{30}$	$L_6, L_3 \rightarrow L_1$	$L_9, L_6 \rightarrow L_8$	-

Ao ser considerada, por exemplo, a previsão para o dia 28 de dezembro de 2001, de acordo com a Tabela 4.1, os índices 5392,43 e 5332,98 dos dias 26 e 27 de dezembro são representados, respectivamente, pelos termos linguísticos L_{48} e L_{47} . Assim, ao ser realizada a busca por RLF com os mesmos antecedentes, encontramos no Grupo 1 uma ocorrência da RLF $L_{48}, L_{47} \rightarrow L_{48}$ e uma ocorrência da RLF $L_{48}, L_{47} \rightarrow$

L_{49} . No Grupo 2, há apenas a RLF $L_{48}, L_{47} \rightarrow L_{48}$ com os antecedentes indicados. No Grupo 3, não há ocorrência de RLF com os antecedentes L_{48} e L_{47} .

Considerando as Equações 4.6 e 4.7, os valores *crisp* que representam os termos linguísticos nas implicações e os pesos $p_{G_1} = 0,22$, $p_{G_2} = 0,33$ e $p_{G_3} = 0,44$, definidos empiricamente para cada grupo, o valor 5421,21 foi previsto.

Em situações onde não há RLF com os antecedentes $L_i = L_{n-1}$ e $L_{i+1} = L_n$, o valor previsto é o parâmetro central do conjunto triangular que representa o termo linguístico L_n .

Após algumas análises da abordagem apresentada nesta seção, dois questionamentos foram levantados. O primeiro é em relação ao número ideal de grupos de RLF. O segundo questionamento é como os graus de influência dos grupos são definidos. A partir de uma análise do modelo de previsão com foco nestes questionamentos, foi possível identificar a possibilidade de combinar o alisamento exponencial simples a uma extensão do modelo de previsão já definido. Esta nova abordagem é apresentada na Seção 4.2.2.

4.2.2 Previsão em STF por meio de uma extensão do alisamento exponencial simples

O modelo de previsão descrito nesta seção considera que as amostras mais recentes da ST devem ter maior influência no cálculo da previsão do que as amostras mais antigas. Entretanto, diferente do modo como ocorre na abordagem da Seção 4.2.1, os conceitos base do AES descrito na Seção 2.3 são combinados com o uso de uma nova proposta para definição de RLF com três termos linguísticos na implicação. O pré-processamento utilizado nesta abordagem de previsão é o modelo apresentado na Seção 4.1.2.

Uma importante característica do modelo de previsão apresentado nesta seção, e não encontrada em outros modelos de STF durante a pesquisa, é a proposta da atualização constante da base de conhecimento por meio da adição de novas RLF. Após o treinamento realizado pelo modelo, onde as RLF são extraídas e a base de conhecimento é definida, a cada nova amostra disponível uma nova RLF é inserida

na base. Os testes realizados, com e sem esta característica nos modelos de previsão, demonstraram uma melhora significativa na acurácia das previsões em cenários onde as ST apresentam, sobretudo no período de previsão, um comportamento diferente do utilizado para treinamento do modelo. Assim, esta abordagem é composta por quatro passos principais.

- 1) Pré-processamento;
- 2) Treinamento do modelo por meio da construção de RLF de segunda ordem com três consequentes;
- 3) Cálculo da previsão a partir da integração das RLF com os conceitos do alisamento exponencial simples;
- 4) Atualização constante da base de conhecimento;

Concluída a definição dos conjuntos fuzzy na fase do pré-processamento, a fuzzificação da ST é realizada para encontrar os termos linguísticos que melhor representem cada um dos valores *crisp* observados na série.

Uma contribuição da abordagem apresentada nesta seção é a extração de RLF de segunda ordem com três termos linguísticos no consequente, onde a suavização é aplicada. Assim, as RLF são extraídas da série temporal fuzzificada como apresentado na Expressão 4.8.

$$L_i, L_{i+1} \rightarrow L_{i+2}, L_{i+3}, L_{i+4}, \quad i = 1, 2, 3, \dots, n - 4. \quad (4.8)$$

onde $1 \leq i \leq n - 4$, n é o número de amostras na série temporal fuzzificada e L_i representa o termo linguístico no instante i . Com três consequentes nas RLF é possível identificar melhor as tendências presentes na ST. Para definir o número mais adequado de consequentes nas RLF, foram realizados diversos experimentos considerando 2 ou mais termos linguísticos na implicação das RLF. Os resultados indicaram que o uso de 3 termos linguísticos no consequente é capaz de representar melhor a tendência comportamental da ST e fornecer uma acurácia satisfatória nas previsões. Com o uso de 4 ou mais consequentes a acurácia das previsões começa a ser prejudicada.

Definida a base de conhecimento, as previsões são realizadas a partir dos dois últimos termos linguísticos que precedem o valor previsto L_{n+1} , ou seja, os termos

L_{n-1} e L_n . Assim, as RLF na base de conhecimento com os antecedentes $L_{n-1} = L_i$ e $L_n = L_{i+1}$ serão utilizadas no cálculo do valor previsto.

Diferente do modo como ocorre no AES clássico, onde a suavização é aplicada a todos os valores da série, nesta abordagem o alisamento é aplicado às RLF. Considerando que as RLF na base de conhecimento apresentam 2 antecedentes e 3 consequentes como apresentado na Expressão 4.8, o primeiro passo é gerar os valores suavizados S a partir de todas as RLF com os antecedentes $L_i = L_{n-1}$ e $L_{i+1} = L_n$, como definido pela Equação 4.9.

$$S = (1 - \alpha_1)^2 l_{i+2} + \alpha_1(1 - \alpha_1) l_{i+3} + \alpha_1 l_{i+4} \quad (4.9)$$

onde α_1 é o índice que indica o grau de suavização, l_{i+2} , l_{i+3} e l_{i+4} representam os valores *crisp* dos termos linguísticos correspondentes, ou seja, o parâmetro central dos respectivos conjuntos fuzzy triangulares e, como no modelo original, a restrição $0 \leq \alpha_1 \leq 1$ deve ser observada. Para atribuir maior influência no cálculo de S ao primeiro termo da implicação representada na Expressão 4.8, α_1 deve ser definido com um valor próximo a 0.

Considere que foram identificadas m RLF na base de conhecimento com os termos linguísticos $L_i = L_{n-1}$ e $L_{i+1} = L_n$ em seus antecedentes. Seja $q = m - 1$ o índice do valor S que representa o comportamento mais recente da ST, ou seja, o valor obtido a partir da Equação 4.9 utilizando a RLF mais recente na base com os antecedentes $L_i = L_{n-1}$ e $L_{i+1} = L_n$. O valor previsto y_{n+1} é calculado como apresentado Equação 4.10.

$$y_{n+1} = \alpha_2 \sum_{i=0}^{q-1} (1 - \alpha_2)^i S_{q-i} + (1 - \alpha_2)^q S_0 \quad (4.10)$$

onde S_0 é o valor calculado pela Equação 4.9 com a RLF mais antiga da base de conhecimento e $0 \leq \alpha_2 \leq 1$ é o fator de suavização. Para atribuir maior influência no cálculo do valor previsto aos valores S que representam o comportamento mais recente da ST, α_2 deve ser definido próximo de 1.

Em situações onde não há ao menos uma RLF com os termos linguísticos $L_i = L_{n-1}$ e $L_{i+1} = L_n$ em seus antecedentes para realizar a previsão, o valor previsto é

definido como o parâmetro central do conjunto triangular que representa o termo linguístico L_n da série. Com o objetivo de diminuir a ocorrência de ocasiões como esta e em busca de melhorar a acurácia nas previsões, o modelo apresentado nesta seção realiza a atualização constante de sua base de conhecimento. Para ilustrar este processo, considere a RLF apresentada na Equação 4.11, onde L_n é o termo linguístico mais recente da STF.

$$L_{n-4}, L_{n-3} \rightarrow L_{n-2}, L_{n-1}, L_n \quad (4.11)$$

A RLF apresentada na Equação 4.11 representa o comportamento mais recente da série, pois é composta pelos últimos 5 termos linguísticos disponíveis. Com a chegada de uma nova amostra, ou seja, do termo linguístico L_{n+1} , o procedimento de atualização adiciona à base de conhecimento do modelo a RLF $L_{n-3}, L_{n-2} \rightarrow L_{n-1}, L_n, L_{n+1}$ contendo o novo termo.

Para ilustrar o procedimento do cálculo de previsão pela abordagem apresentada nesta seção, considere a os índices da série temporal TAIEX 2003, utilizada nos experimentos apresentados no Capítulo 6. Após o pré-processamento utilizando a abordagem apresentada na Seção 4.1.2, os termos linguísticos obtidos para representação da ST são apresentados na Tabela 4.5.

Tabela 4.5 – Valores *crisp* de representação para os termos linguísticos do índice TAIEX 2003

Termo linguístico	Valor	Termo linguístico	Valor	Termo linguístico	Valor	Termo linguístico	Valor
0	4439,08	13	5223,69	26	5488,22	39	5853,37
1	4537,30	14	5257,34	27	5513,87	40	5869,14
2	4590,80	15	5284,78	28	5522,88	41	5917,75
3	4648,07	16	5301,02	29	5553,36	42	5925,46
4	4691,68	17	5303,11	30	5582,89	43	5957,99
5	4827,69	18	5318,04	31	5620,80	44	6038,06
6	4893,06	19	5341,17	32	5645,91	45	6039,01
7	4939,72	20	5367,29	33	5678,97	46	6044,48
8	4970,38	21	5393,38	34	5694,98	47	6066,15
9	4997,74	22	5409,20	35	5721,49	48	6094,29
10	5075,14	23	5438,26	36	5750,26	-	-
11	5144,70	24	5448,49	37	5817,84	-	-
12	5203,78	25	5486,13	38	5818,87	-	-

Definidos os termos linguísticos, a fuzzificação da ST é realizada. Considerando os termos listados na Tabela 4.5 e os índices TAIEX observados entre janeiro e outubro de 2003, a série temporal fuzzificada utilizada para treinamento do modelo é apresentada na Tabela 4.6.

Tabela 4.6 – Índices TAIEX 2003 fuzzificados para treinamento do modelo

$L_1, L_3, L_4, L_4, L_5, L_5, L_5, L_9, L_9, L_9, L_7, L_6, L_7, L_7, L_9, L_{10}, L_{10}, L_8, L_9, L_5, L_4, L_3, L_2, L_3, L_1, L_1,$ $L_4, L_2, L_1, L_1, L_1, L_2, L_0, L_0, L_0, L_1, L_1, L_0, L_0, L_0, L_0, L_0, L_0, L_0, L_0, L_1, L_1, L_2, L_2, L_2, L_1, L_1,$ $L_1, L_0, L_0, L_0, L_0, L_0, L_1, L_2, L_1, L_1, L_1, L_1, L_0, L_0, L_2, L_2, L_3, L_3, L_1, L_2, L_0, L_0, L_0, L_0, L_0, L_0,$ $L_0, L_0, L_0, L_0, L_0, L_0, L_0, L_0, L_0, L_0, L_0, L_0, L_0, L_0, L_0, L_0, L_1, L_1, L_4, L_4, L_4, L_4, L_5, L_5, L_5, L_6,$ $L_6, L_6, L_8, L_9, L_{10}, L_9, L_7, L_6, L_7, L_6, L_6, L_6, L_9, L_{10}, L_{10}, L_{11}, L_{18}, L_{20}, L_{20}, L_{15}, L_{13}, L_{19}, L_{19}, L_{22},$ $L_{16}, L_{15}, L_{13}, L_{14}, L_{15}, L_{22}, L_{21}, L_{24}, L_{19}, L_{17}, L_{18}, L_{21}, L_{20}, L_{14}, L_{13}, L_{14}, L_{13}, L_{13}, L_{14}, L_{23}, L_{23},$ $L_{26}, L_{27}, L_{29}, L_{31}, L_{32}, L_{33}, L_{29}, L_{29}, L_{28}, L_{32}, L_{34}, L_{35}, L_{33}, L_{31}, L_{32}, L_{35}, L_{33}, L_{31}, L_{32},$ $L_{31}, L_{34}, L_{36}, L_{36}, L_{36}, L_{33}, L_{33}, L_{35}, L_{34}, L_{32}, L_{32}, L_{31}, L_{30}, L_{34}, L_{36}, L_{39}, L_{39}, L_{38},$ $L_{40}, L_{43}, L_{42}, L_{42}, L_{44}, L_{46}, L_{47}, L_{47}, L_{46}, L_{43}, L_{41}, L_{43}, L_{47}, L_{48}, L_{48}.$

O próximo passo definido pelo modelo é extrair as RLF de segunda ordem com três consequentes, como apresentado na Expressão 4.8. De acordo com a Tabela 4.6, a primeira RLF a compor a base de conhecimento é $L_1, L_3 \rightarrow L_4, L_4, L_5$, seguida pela RLF $L_3, L_4 \rightarrow L_4, L_5, L_5$ e assim por diante.

Com o objetivo de comprovar a importância da atualização constante da base de conhecimento com a inclusão de novas RLF, considere a previsão para o dia 29 de dezembro de 2003 com o valor real do índice TAIEX igual a 5804,89. Os dois termos linguísticos L_{n-1} e L_n , que antecedem a previsão, são representados respectivamente por L_{39} e L_{39} . Na ST fuzzificada apresentada na Tabela 4.6, é possível identificar somente a RLF $L_{39}, L_{39} \rightarrow L_{38}, L_{40}, L_{43}$ para cálculo do valor S como apresentado na Equação 4.9. Considerando $\alpha_1 = 0,1$, os pesos dos 3 consequentes da RLF a partir do primeiro termo na implicação são, respectivamente, 0,81, 0,09 e 0,1. A partir dos valores *crisp* para representação dos termos linguísticos apresentados na Tabela 4.5, e considerando que há somente uma RLF para o cálculo do valor previsto, o resultado da previsão seria 5837,30. Entretanto, depois do treinamento inicial do modelo e antes da data de 29 de dezembro de 2003, a base de conhecimento foi atualizada com novas RLF definidas a partir da chegada de novas amostras na série. Com a sequência obtida no intervalo de 25 de novembro à 1º de dezembro, a RLF $L_{39}, L_{39} \rightarrow L_{36}, L_{36}, L_{40}$ foi inserida na base de conhecimento e, os dados observados entre 8 e 12 de dezembro, permitiram a inclusão da RLF $L_{39}, L_{39} \rightarrow L_{37}, L_{40}, L_{39}$.

A partir da base de conhecimento atualizada e considerando $\alpha_1 = 0,1$ para o cálculo dos valores S na Equação 4.9, foram definidos $S_0 = 5837,30$, $S_1 = 5762,14$ e $S_2 = 5826,01$, sendo S_2 o valor obtido pela RLF inserida mais recentemente na base. No cálculo da Equação 4.10, o valor atribuído à α_2 foi 0,5, o que resultou na previsão do valor 5812,87, mais próximo do valor real da série do que a previsão anterior sem a atualização da base.

Embora os resultados tenham apresentado uma evolução na acurácia das previsões em comparação com a abordagem anterior, o modelo de alisamento apresentado nas Equações 4.9 e 4.10 ainda necessita de alguns ajustes. Ao utilizar $\alpha_1 = 0,5$ em uma sequência de três valores, temos para o primeiro e segundo termos o peso 0,25 e para o terceiro termo o peso de 0,5. Quando $\alpha_1 = 0,8$ temos para o primeiro termo o fator de ponderação 0,04, para o segundo termo 0,16 e para o terceiro termo 0,8. Como é possível observar para $0,5 < \alpha < 1$, os graus de influência apresentam um comportamento coerente em relação ao aumento dos fatores de ponderação. Entretanto, quando $\alpha_1 < 0,5$ a diminuição dos valores não ocorre como esperado, pois o terceiro termo sempre terá um grau de influência maior que o do segundo, podendo em alguns casos, ser atribuído ao terceiro termo um fator de ponderação maior que o definido para o primeiro, como ocorre, por exemplo, quando $\alpha_1 = 0,45$.

Considerando as limitações do modelo atual, uma nova técnica de alisamento para suavização dos valores foi desenvolvida. Com o objetivo de auxiliar nas previsões em ocasiões onde não há RLF na base de conhecimento para os cálculos da previsão, os conceitos de regressão linear também foram introduzidos ao modelo. Esta nova abordagem é apresentada na Seção 4.2.3.

4.2.3 Um modelo híbrido de previsão de valores combinando séries temporais fuzzy, regressão linear e uma nova técnica de suavização

Diferente das outras duas abordagens de previsão, o modelo de previsão apresentado nesta seção utiliza como técnica de pré-processamento a abordagem descrita na Seção 4.1.3. São duas as principais contribuições do modelo de previsão descrito nesta seção. A primeira delas é a introdução de uma nova técnica de suavização para ser aplicada às RLF. A segunda contribuição é o uso da regressão linear simples para

auxiliar nas previsões onde, mesmo com a atualização constante da base de conhecimento, não há ao menos uma RLF com os termos linguísticos $L_i = L_{n-1}$ e $L_{i+1} = L_n$.

De modo similar ao descrito na Seção 4.2.2, a abordagem de previsão apresentada nesta seção também realiza a suavização dos consequentes das RLF com antecedentes $L_i = L_{n-1}$ e $L_{i+1} = L_n$. Entretanto, os índices de suavização são definidos conforme apresentado na Equação 4.12.

$$w_i = \frac{\theta^i}{\sum_{j=0}^{n-1} \theta^j} \quad (4.12)$$

onde w_i é o grau de influência suavizado do termo i , n é o número de termos a serem suavizados e θ é o fator de suavização. Como é possível observar na Equação 4.12, o índice da primeira amostra a ser suavizada deve ser definido como 0. Se $0 < \theta < 1$, o grau de influência dos termos próximos a n terão maior influência no cálculo do valor suavizado. Se $\theta = 0$, somente o primeiro termo da série será considerado. Quando $\theta = 1$, todos os termos terão o mesmo grau de influência. Se $\theta > 1$, os primeiros termos disponíveis na série terão maior influência no cálculo do valor suavizado. Para ilustrar o comportamento da técnica de suavização proposta, a Figura 4.4 apresenta os graus de suavização para 3 amostras considerando $\theta < 1$.

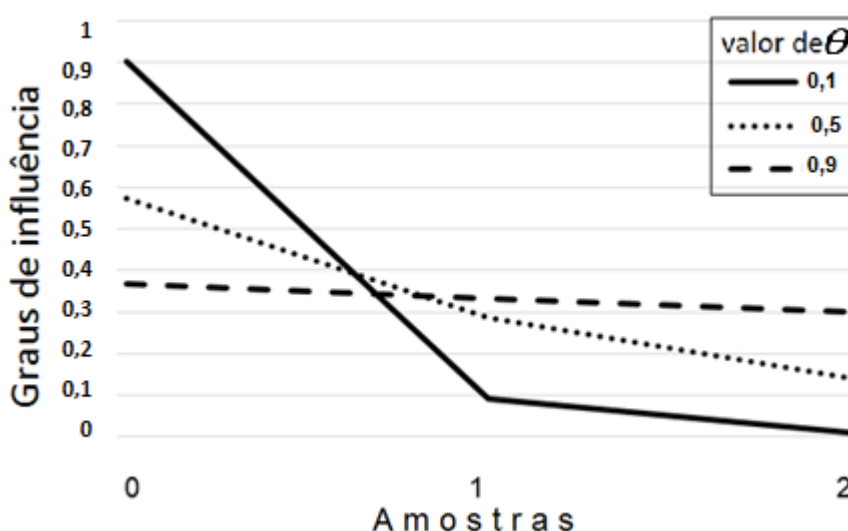


Figura 4.4 – Comportamento dos graus de influência suavizados para $\theta < 1$ com 3 amostras

No gráfico apresentado pela Figura 4.4 é possível observar que a suavização dos graus de influência diminui drasticamente quando θ está próximo a 0. Se o valor

de θ for definido próximo a 1, a queda nos graus de influência das amostras ocorre lentamente. A Figura 4.5 apresenta o comportamento dos graus de influência para $\theta > 1$, considerando 3 termos na suavização.

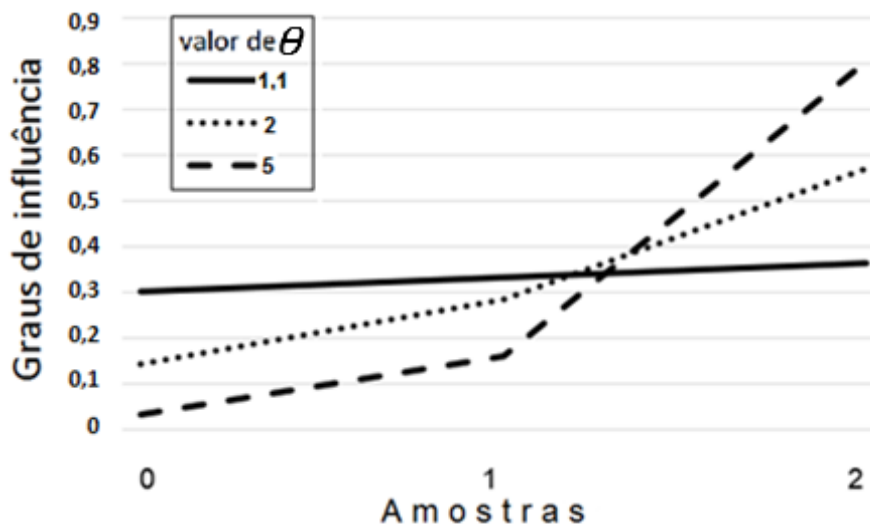


Figura 4.5 - Comportamento dos graus de influência suavizados para $\theta > 1$ com 3 amostras

Na Figura 4.5 é possível observar que, quando θ é definido distante de 1, os graus de influência no cálculo do valor suavizado aumentam drasticamente. Entretanto, quando θ está próximo de 0, os graus de influência dos termos suavizados aumentam lentamente.

A partir das amostras fuzzificadas e da definição da base de conhecimento inicial, composta por RLF de segunda ordem com três consequentes, o primeiro passo do processo de previsão é buscar por RLF na base de conhecimento. Identificadas as RLF com os antecedentes $L_i = L_{n-1}$ e $L_{i+1} = L_n$, o próximo passo é aplicar a suavização nos consequentes destas relações. Por fim, os valores obtidos são suavizados novamente para a definição do valor previsto. Caso não haja ao menos uma RLF na base de conhecimento o cálculo do valor previsto, o processo de previsão é realizado com o auxílio da regressão linear simples. A cada nova amostra disponível, a base de conhecimento é atualizada.

Considerando os 3 consequentes de cada RLF com os termos linguísticos $L_i = L_{n-1}$ e $L_{i+1} = L_n$, o cálculo dos valores suavizados S de cada RLF é realizado como apresentado na Equação 4.13.

$$S = \sum_{i=0}^{n-1} w_i \cdot l_i \quad (4.13)$$

onde $n = 3$ indica o número de elementos no consequente das RLF, l_i é valor *crisp* que representa o termo linguístico i , ou seja, o parâmetro central do conjunto triangular L_i , w_i é definido pela Equação 4.12 e S é o valor suavizado.

Considere que foram identificadas m RLF na base de conhecimento com os termos linguísticos $L_i = L_{n-1}$ e $L_{i+1} = L_n$. Os m valores definidos pela Equação 4.13 irão compor uma nova série, onde o valor S_0 , inferido a partir da RLF mais antiga utilizada para o cálculo dos valores S , será a primeira amostra da nova série. De modo análogo, a RLF mais recente utilizada no cálculo do valor S_{m-1} será a última amostra na nova série. O valor previsto y_{n+1} é definido a partir de uma nova suavização aplicada à sequência dos r valores calculados a partir da Equação 4.13. Este cálculo é apresentado na Equação 4.14.

$$y_{n+1} = \sum_{i=0}^{r-1} w_i \cdot S_i \quad (4.14)$$

onde w_i é definido pela Equação 4.12, S_i é o parâmetro central do conjunto fuzzy que melhor representa o valor suavizado calculado pela Equação 4.13 a partir das r RLF com os antecedentes $L_i = L_{n-1}$ e $L_{i+1} = L_n$ e y_{n+1} é o valor suavizado a ser utilizado na previsão.

Diversos testes foram realizados para analisar a acurácia dos valores previstos em uma ST comparando o valor fornecido pela Equação 4.14, com o parâmetro central do conjunto fuzzy que melhor representa y_{n+1} . Os testes indicaram que, embora haja uma redução no número de previsões exatas, uma parcela significativa das previsões por meio do parâmetro central dos conjuntos fuzzy triangulares fica próxima dos valores observados, muitas vezes gerando taxas de erro irrelevantes, o que acaba aumentando a acurácia geral das previsões. Desta forma, foi possível concluir que a melhor alternativa para o modelo é gerar previsões muito próximas dos valores observados na grande maioria das vezes, do que optar por poucas previsões exatas com algumas distorções que acabam prejudicando a acurácia geral do modelo.

Outra importante contribuição desta abordagem, é a combinação de regressão linear com os conceitos de STF. Ao invés de utilizar o valor do parâmetro central do conjunto triangular que representa o termo linguístico L_n em ocasiões onde não há ao menos uma RLF com os antecedentes $L_i = L_{n-1}$ e $L_{i+1} = L_n$, nesta abordagem é utilizado o parâmetro $\hat{\beta}_1$ do modelo de regressão linear que indica a variação em Y_t para cada aumento de uma unidade em X_t . Assim, quando não há RLF disponíveis para realizar a previsão por meio das Equações 4.13 e 4.14, o valor previsto é definido como apresentado na Equação 4.15.

$$y_{n+1} = l_n + \hat{\beta}_1 \quad (4.15)$$

onde y_{n+1} é o valor previsto, l_n é o valor *crisp* do parâmetro central do conjunto fuzzy triangular que representa o termo linguístico mais recente da série, e $\hat{\beta}_1$ é o parâmetro calculado pela Equação 2.4. Com o auxílio da regressão linear simples nas previsões como apresentado neste modelo, é possível obter uma melhora na acurácia agregando a tendência de variação da série ao valor da última amostra disponível.

Um parâmetro importante para o cálculo de $\hat{\beta}_1$ é o número de amostras a serem consideradas. Ao considerarmos, por exemplo, a ST desde a sua primeira amostra, o valor de $\hat{\beta}_1$ pode apresentar uma variação que não represente adequadamente o comportamento atual da série. Por outro lado, ao ser considerado um número reduzido de amostras em seu cálculo, esta quantidade de amostras pode não ser suficiente para identificar, satisfatoriamente, a variação média real do comportamento recente da série. A definição adequada deste parâmetro ainda sofre influência do tamanho da ST e do comportamento apresentado por ela.

Após a análise de diversos experimentos, foram obtidos resultados satisfatórios utilizando um número aproximado de 20% (vinte por cento) do total de amostras disponíveis, ou seja, em uma ST com 50 amostras por exemplo, os 10 valores mais recentes são utilizados no cálculo de $\hat{\beta}_1$.

Para demonstrar o passo a passo no cálculo do valor previsto pelo modelo de previsão apresentado nesta seção, considere os termos linguísticos dos índices TAIX 2002 com seus respectivos valores *crisp* de representação, apresentados na Tabela 4.7.

Tabela 4.7 – Valores *crisp* de representação para os termos linguísticos do índice TAIEX 2002

Termo lin- guístico	Valor	Termo lin- guístico	Valor	Termo lin- guístico	Valor	Termo lin- guístico	Valor
0	3209,20	22	4222,81	44	4936,69	66	6091,94
1	3254,98	23	4280,93	45	4966,86	67	6276,42
2	3300,76	24	4287,06	46	5000,35	68	6278,42
3	3346,54	25	4328,56	47	5044,12	69	6324,20
4	3392,32	26	4386,80	48	5072,30	70	6369,97
5	3438,10	27	4430,67	49	5127,79	71	6415,75
6	3483,88	28	4463,98	50	5160,66	72	6461,53
7	3529,66	29	4531,21	51	5251,76	73	6507,31
8	3575,44	30	4560,86	52	5317,39	74	6553,09
9	3621,22	31	4579,39	53	5394,30	75	6598,87
10	3666,99	32	4598,21	54	5443,52	76	6644,65
11	3712,77	33	4635,28	55	5499,93	77	6690,43
12	3758,55	34	4646,03	56	5533,52	78	6736,21
13	3804,33	35	4660,64	57	5566,81	79	6781,99
14	3850,11	36	4668,14	58	5599,27	80	6827,77
15	3917,56	37	4700,41	59	5631,98	81	6873,55
16	3956,42	38	4720,96	60	5666,98	82	6919,33
17	4071,91	39	4766,15	61	5696,33	83	6965,11
18	4131,56	40	4801,85	62	5743,14	84	7010,89
19	4167,11	41	4854,20	63	5805,63	85	7056,66
20	4188,86	42	4886,03	64	5867,94	-	-
21	4208,86	43	4919,11	65	5966,14	-	-

Após a fuzzificação da ST e a extração das RLF com 2 antecedentes e 3 consequentes, os cálculos para previsão são realizados. Quando são encontradas uma ou mais RLF na base de conhecimento com os antecedentes $L_i = L_{n-1}$ e $L_{i+1} = L_n$, os valores *crisp* de representação dos termos linguísticos dos consequentes são suavizados pelas Equações 4.12 e 4.13. Cada valor S obtido irá compor um vetor que será utilizado para a definição do valor previsto por meio da Equação 4.14. O valor *crisp* do termo linguístico que melhor representar y_{n+1} , será a previsão fornecida pelo modelo.

Considere a previsão do índice TAIEX para o dia 18 de dezembro de 2002. Os 2 termos linguísticos mais recentes disponíveis na série, referentes aos dias 16 e 17

de dezembro são, respectivamente, L_{31} e L_{29} . Considerando estes 2 termos linguísticos, duas RLF serão utilizadas para a previsão do dia 18 de dezembro de 2002, $L_{31}, L_{29} \rightarrow L_{28}, L_{29}, L_{36}$ e $L_{31}, L_{29} \rightarrow L_{31}, L_{30}, L_{38}$. Ao considerarmos $\theta = 0,4$, os fatores de suavização a serem aplicados nos consequentes das RLF são 0,64, 0,26 e 0,1, respectivamente, para o primeiro, o segundo e o terceiro termos linguísticos. Assim, a partir da Equação 4.13 temos $S_0 = 4502,15$ e $S_1 = 4589,15$. De acordo com a Equação 4.14, e considerando $\theta = 0,2$ para a definição de w_i , temos $y_{n+1} = 4542,37$. Assim, ao considerarmos o parâmetro central do conjunto triangular fuzzy com maior pertinência para o resultado, temos o valor 4531,21 como índice previsto.

Entretanto, quando não há ao menos uma RLF na base de conhecimento para o cálculo da previsão, a predição deve ser realizada com o auxílio da regressão linear simples. Considere a previsão do índice TAIEX para 25 de novembro de 2002, com os antecedentes L_{31} e L_{37} , situação onde não há ao menos uma RLF para o cálculo. Ao serem consideradas as 41 amostras mais recentes disponíveis na STF, temos $\hat{\beta}_1 = 20,435$, conforme apresentado no gráfico da Figura 4.6.

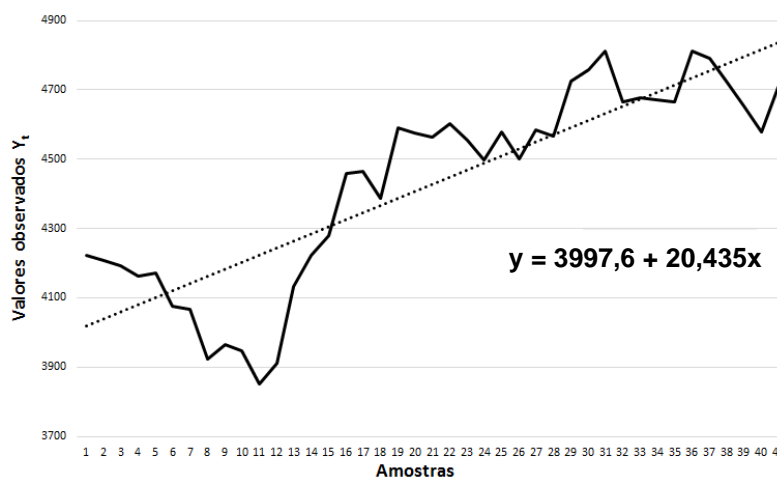


Figura 4.6 – Reta de regressão para a previsão do índice TAIEX de 25 de novembro de 2002

De acordo com a Tabela 4.7, ao ser considerado o valor 4700,41 que representa o parâmetro central do termo linguístico L_{37} , ou seja, o termo linguístico mais recente disponível na STF, e o valor de $\hat{\beta}_1 = 20,435$ ilustrado na Figura 4.6, de acordo com a Equação 4.15 temos $y_{n+1} = 4720,845$ com L_{38} como o termo linguístico que

melhor representa y_{n+1} sendo, portanto, 4720,96 o valor previsto e 4723,16 o valor real.

4.3 Considerações finais

Além das abordagens desenvolvidas para o pré-processamento e para a previsão de valores, neste capítulo também foi demonstrada a evolução dos modelos ao longo do desenvolvimento da pesquisa. Os ajustes realizados no pré-processamento, a suavização aplicada às RLF e o uso da regressão linear como apoio no cálculo dos valores previstos, certamente são os principais responsáveis pela eficácia do modelo de previsão definido.

No próximo capítulo é apresentada a abordagem desenvolvida para a análise do comportamento dinâmico das séries temporais onde um novo modelo para o cálculo da dissimilaridade entre os segmentos das STF é apresentado. Este novo modelo tem como base para a análise comportamental os conceitos da métrica DTW aplicados na análise segmentada de uma ST combinados com uma nova técnica de ponderação das células da matriz DTW.

Capítulo 5

ANÁLISE DO COMPORTAMENTO DINÂMICO DAS STF

Após uma análise detalhada das abordagens para agrupamento de ST disponíveis na literatura, foi definido como objetivo do modelo realizar o agrupamento das ST considerando a possibilidade de mudança comportamental de uma série ao longo do período de observação.

Assim como nos modelos de previsão, na análise comportamental das séries temporais também é realizado o pré-processamento. O modelo utilizado é o apresentado na Seção 4.1.3 sem a remoção de *outliers*. Entretanto, ao invés de ser considerada apenas uma ST durante a análise, todas as séries da base são analisadas. Para a definição do domínio, todas as amostras da base são consideradas, ou seja, se houver 100 ST com 50 amostras cada uma, serão consideradas as 5000 amostras na definição do domínio como se fossem uma série única. Entretanto, na definição do número de conjuntos, as ST são avaliadas uma a uma. O maior número de termos linguísticos indicado por uma ST será utilizado como parâmetro para a definição dos termos linguísticos. As etapas seguintes para a análise comportamental das ST são apresentadas na Figura 5.1.



Figura 5.1 – Etapas do processo de análise comportamental das STF

Para realizar uma análise do comportamento dinâmico de uma STF, de acordo com a abordagem desenvolvida, o primeiro passo após o pré-processamento dos dados é a definição dos protótipos a partir das séries temporais disponíveis para treinamento. Os protótipos tem como objetivo representar, do melhor modo possível, o comportamento das STF pertencentes à cada grupo.

Definidos os protótipos, as STF presentes na base de testes devem ser segmentadas. Este procedimento é necessário pois a análise comportamental das séries é realizada considerando segmentos isolados ao longo do período de observação. Por reduzir consideravelmente o espaço de busca em agrupamentos com STF longas, a segmentação também contribui significativamente para a redução do custo computacional.

Definidos os segmentos de cada STF na base de testes, uma extensão do algoritmo DTW ponderado aplicado aos conjuntos fuzzy é utilizada para definir a similaridade entre um determinado segmento de uma STF, e o intervalo de observação correspondente dos protótipos já definidos. A ponderação é realizada por meio de uma nova técnica de suavização que permite um maior controle dos fatores de ponderação utilizados.

O restante deste capítulo apresenta como os protótipos são definidos, os procedimentos utilizados para a segmentação das séries temporais e, por fim, como a similaridade entre as séries é definida.

5.1 Definição dos protótipos

Com o objetivo de buscar a melhor representação para um conjunto de STF por meio da definição de um protótipo, diversos experimentos foram realizados com diferentes métricas para avaliar a similaridade entre as séries temporais. Com base nos resultados obtidos nestes experimentos, o algoritmo Fuzzy C-Means (BEZDEK, 1981)

combinado com a métrica DTW clássica (BELLMAN, 1957) foi o modelo escolhido para a definição dos protótipos que representarão o comportamento das ST pertencentes a cada grupo. Entretanto, experimentos utilizando outras métricas como a distância euclidiana e algumas variações da métrica DTW, também foram realizados, mas sem resultados significativos em relação a definição dos protótipos. Também foram realizados experimentos onde o protótipo foi definido por meio da média e da mediana dos valores observados para cada instante i da série. Entretanto, os protótipos obtidos não eram capazes de representar adequadamente os dados observados.

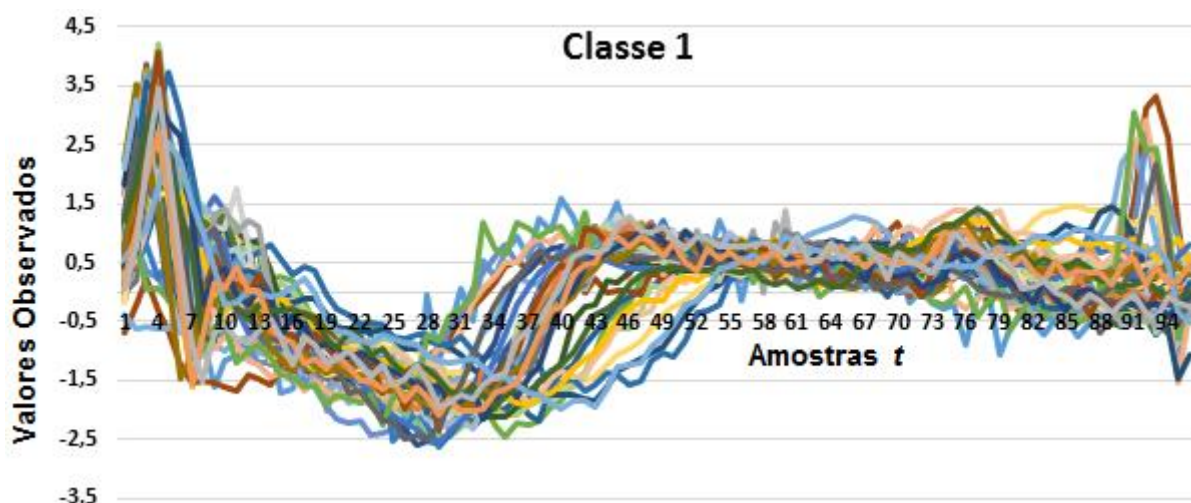
O primeiro passo para a definição dos protótipos é identificar se há ST rotuladas na base de treinamento, ou seja, verificar se as ST possuem rótulos que indiquem a qual grupo cada série pertence. Em problemas onde há dados rotulados disponíveis, a matriz de pertinência do FCM (BEZDEK, 1981) é inicializada com cada ST tendo um grau de pertinência maior que 50% (cinquenta por cento) no *cluster* de seu rótulo, e o restante distribuído nos demais *clusters*. Em problemas onde não há dados rotulados, o número de *clusters* é definido a partir de uma análise da representação gráfica das séries, e a inicialização da matriz de pertinência é feita de forma aleatória.

Definidos o número de grupos e concluída a inicialização da matriz de pertinência, o algoritmo FCM (BEZDEK, 1981) é executado até que o número máximo de iterações seja atingido, ou enquanto não for satisfeita a condição do ganho mínimo δ entre duas iterações seguidas.

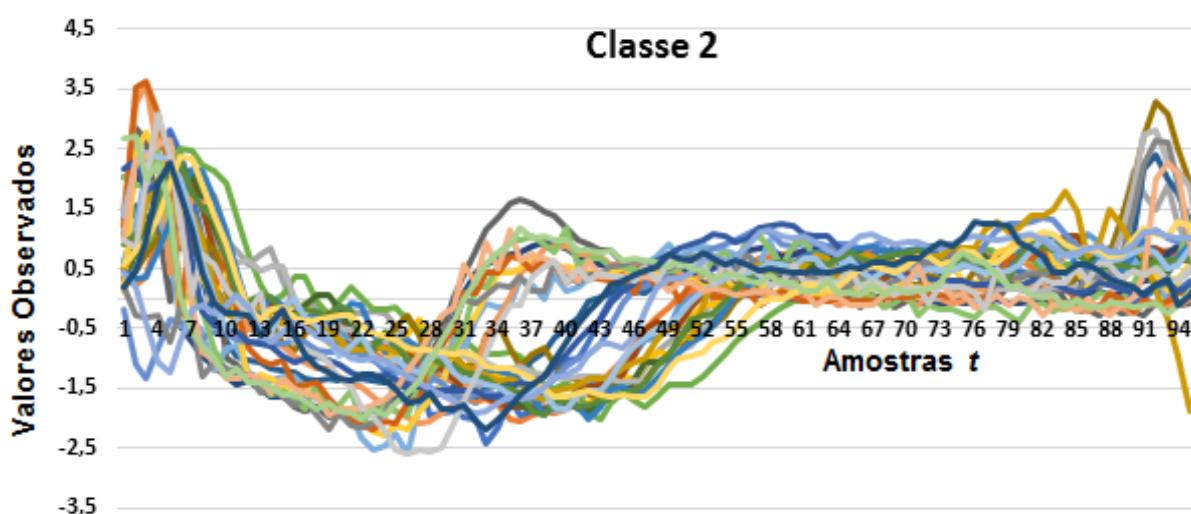
Para demonstrar a abordagem desenvolvida, considere a base de dados ECG200 (Eletrocardiograma 200), obtida da base de dados da *University of California, Riverside* (UCR) (KEOGH, et al., 2011), com 200 séries temporais contendo 96 amostras cada uma. De acordo com a distribuição dos dados sugerida em (KEOGH, et al., 2011), a base de treinamento é definida com 100 séries temporais rotuladas divididas em 2 classes. A Figura 5.2 apresenta a representação gráfica das séries rotuladas da Classe 1 e da Classe 2.

É possível perceber em uma análise visual que, predominantemente, as ST da Classe 1 apresentam um comportamento mais agudo no início da série, ao passo que as séries da Classe 2 apresentam uma curva um pouco mais acentuada antes da queda nos valores observados em ambas as classes. Na Classe 1, a maior parte das ST voltam a apresentar elevação nos valores entre as amostras 34 e 40, enquanto na Classe 2 esta alteração ocorre, em sua maioria, de uma forma mais acentuada entre

as amostras 49 e 54. Durante as últimas amostras das séries temporais da Classe 1, os valores voltam a apresentar um comportamento um pouco mais agudo quando comparado com o mesmo período de observações das séries temporais da Classe 2.



(a) Sessenta e nove ST da classe 1



(b) Trinta e uma ST da classe 2

Figura 5.2 – Séries temporais da base ECG200 utilizadas para a definição dos protótipos (KEOGH, et al., 2011)

Para ilustrar a definição dos protótipos por meio do FCM (BEZDEK, 1981) combinado com a métrica DTW (BELLMAN, 1957) e também visando permitir uma análise comparativa dos resultados, a Figura 5.3 apresenta os protótipos das séries temporais obtidos por meio do FCM clássico, utilizando a distância euclidiana como métrica de dissimilaridade.

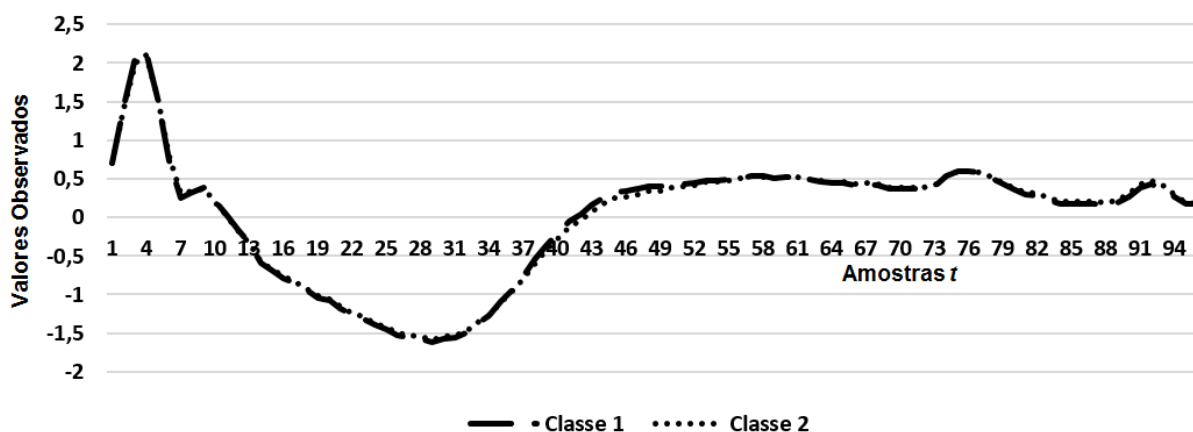


Figura 5.3 – Protótipos das séries temporais definidos por meio do Fuzzy C-Means clássico utilizando a distância euclidiana

A Figura 5.4 apresenta os protótipos obtidos por meio do Fuzzy C-Means (BEZDEK, 1981) utilizando a métrica DTW (BELLMAN, 1957) clássica para a análise da dissimilaridade das séries.

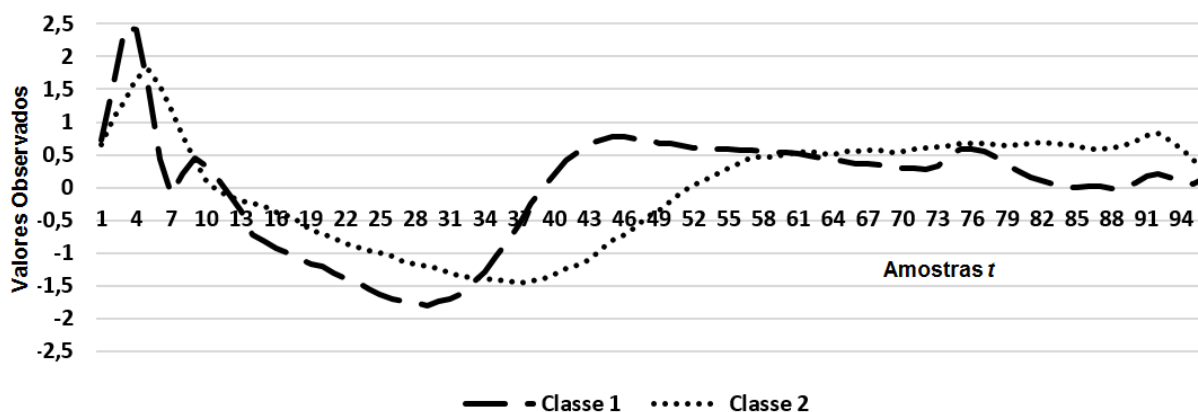


Figura 5.4 – Protótipos das séries temporais definidos por meio do Fuzzy C-Means utilizando a métrica DTW

Ambos os experimentos das Figuras 5.3 e 5.4 foram executados com os mesmos parâmetros do Fuzzy C-Means, alterando apenas a métrica para calcular a dissimilaridade entre as séries temporais. O resultado apresentado na Figura 5.3 comprova o problema introduzido na Seção 3.1.1 e ilustrado na Figura 3.2 em relação ao uso da distância euclidiana para análise de séries temporais. Por outro lado, em razão da flexibilidade presente na análise na métrica DTW, os protótipos da Figura 5.4 representam mais adequadamente os dois grupos de séries temporais apresentados da Figura 5.2.

5.2 Segmentação das séries temporais

Para realizar a análise segmentada das ST e permitir o agrupamento dinâmico das mesmas, a abordagem desenvolvida utiliza os conceitos de regressão linear simples para a definição de R^2 , como definido pela Equação 2.7, combinado com o algoritmo *bottom-up* apresentado na Seção 2.4. O objetivo deste processo é particionar a ST para que o comportamento de cada segmento seja avaliado isoladamente do restante da série.

Após a estimativa dos parâmetros $\hat{\beta}_0$ e $\hat{\beta}_1$, respectivamente, pelas Equações 2.3 e 2.4, o modelo estimado pela regressão linear simples que representa os dados observados é definido pela Equação 5.1.

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_t \quad (5.1)$$

onde, \hat{y}_t é o valor estimado para a variável Y_t no instante t e x é o valor observado na variável X_t .

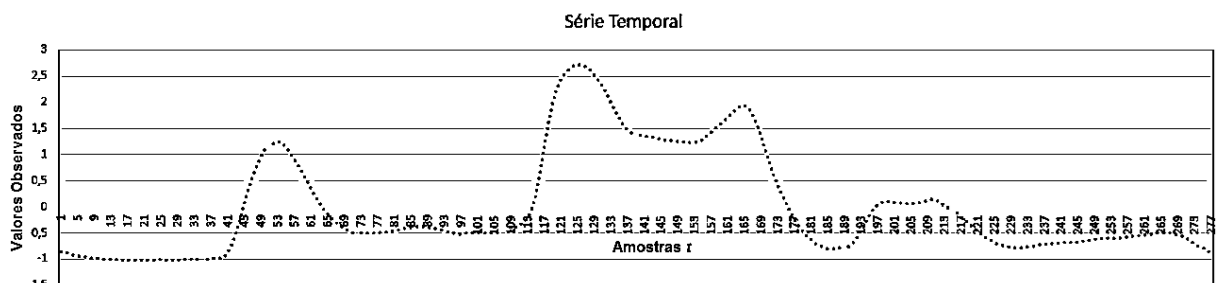
Considerando que o valor de R^2 indica, em outras palavras, o quão linearmente estão organizados os dados observados ao longo do tempo e que, como constatado em Keogh, et al., (2001), o algoritmo *bottom-up* apresenta os melhores resultados dentre as três principais técnicas para segmentação, a abordagem desenvolvida inicialmente considera $n/2$ segmentos ao longo da série, onde n é o número de amostras disponíveis na série.

Os segmentos iniciais são definidos pela união de duas amostras vizinhas realizada ao longo de toda a série temporal a partir da primeira amostra, ou seja, a primeira amostra irá compor o segmento 1 juntamente com a segunda amostra. O terceiro e o quarto dados disponíveis na série formam o segmento 2, e assim por diante. Caso a ST tenha um número ímpar de amostras, o último segmento a ser definido neste processo será composto de três elementos.

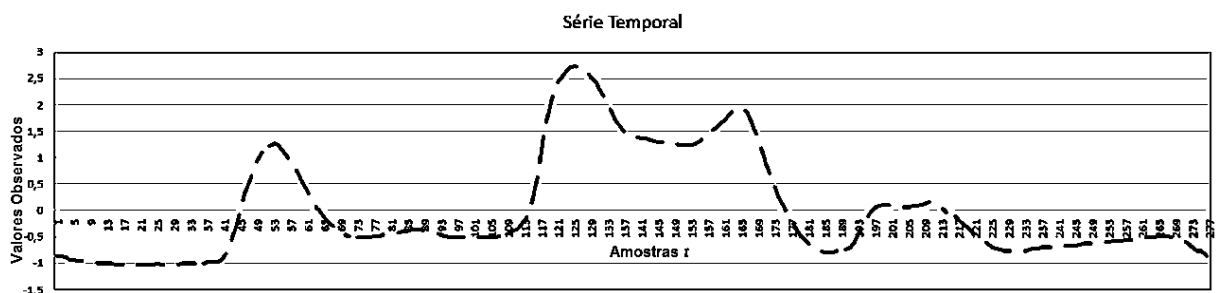
Como os segmentos definidos anteriormente possuem apenas duas amostras, com possível exceção ao último segmento da série, o índice R^2 tem seu valor igual a 1. Assim, o próximo passo é realizar a união dos dois segmentos vizinhos que resultam no maior valor de R^2 , ou seja, o índice é calculado para a união do segmento 1

com o segmento 2, do segmento 2 com o segmento 3, e assim por diante. Se o maior valor de R^2 obtido ao longo de toda a série for maior que um *threshold* definido pelo usuário, a união dos respectivos segmentos é realizada.

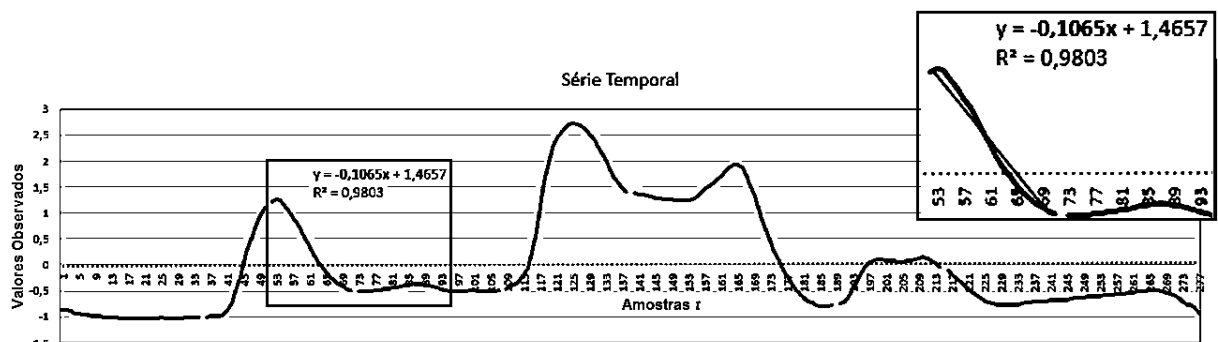
Considerando que a partir da união de dois segmentos vizinhos um novo segmento é definido, um novo valor de R^2 deve ser calculado considerando a possível união do novo segmento com seus respectivos vizinhos. Enquanto o maior valor de R^2 ao longo de toda a série for maior que o *threshold* definido pelo usuário, este procedimento é repetido. A Figura 5.5 apresenta a evolução do processo de segmentação conforme a abordagem descrita nesta seção.



(a) Segmentos iniciais



(b) Série temporal segmentada após n iterações



(c) Série temporal segmentada ao final do processo

Figura 5.5 – Evolução do processo de segmentação conforme abordagem apresentada

Um parâmetro importante na segmentação da série é o *threshold* definido pelo usuário. Se for definido um valor muito baixo para R^2 , próximo de 0,3 por exemplo, a ST será segmentada em um número menor de segmentos com um número maior de amostras em cada um. Entretanto, se a ST apresentar um comportamento irregular ou mesmo um comportamento com tendências de crescimento e diminuição de valores intercaladas, possivelmente, dentro de um mesmo segmento haverá intervalos com diferentes comportamentos, situação que deve ser evitada de acordo com a abordagem apresentada. Por outro lado, a definição de um *threshold* acima de 0,98 por exemplo, pode resultar em segmentos compostos por um número muito reduzido de amostras, o que acaba não permitindo a identificação comportamental do período analisado.

No detalhe da Figura 5.5(c) é possível visualizar o modelo estimado pela regressão linear ao segmento em destaque e seu valor de R^2 . O valor mínimo definido para R^2 na Figura 5.5 foi 0,9. Experimentos em diferentes ST indicam valores entre 0,8 e 0,95 como os mais indicados para a definição do *threshold*.

5.3 Agrupamento dinâmico das séries temporais

Com a definição dos protótipos e a segmentação das ST concluída, a dissimilaridade entre um segmento de uma ST e o intervalo equivalente de cada protótipo é definida por meio de uma nova proposta de extensão do algoritmo DTW para, na sequência, ser calculada a pertinência do segmento nos respectivos grupos. Este processo é apresentado nas Seções 5.3.1 e 5.3.2.

5.3.1 Algoritmo DTW fuzzy ponderado para cálculo da dissimilaridade

Para definir a dissimilaridade de um segmento em relação ao respectivo intervalo de um determinado protótipo e, posteriormente, calcular a pertinência deste segmento ao grupo que o protótipo representa, o primeiro passo é preencher a matriz DTW com as distâncias entre os dados observados como apresentado na Figura 3.4. Entretanto, a dissimilaridade calculada entre as amostras e representada em cada

célula da matriz, deve ser calculada a partir dos dados fuzzificados e, portanto, considerando os conjuntos fuzzy triangulares que representam os dados observados. A diferença entre os conjuntos fuzzy triangulares A e B , que representam os dados fuzzificados de um segmento e de um protótipo, é calculada como apresentado na Equação 5.2.

$$d(A, B) = \sqrt{\frac{1}{3} [(a_1 - b_1)^2 + (a_2 - b_2)^2 + (a_3 - b_3)^2]} \quad (5.2)$$

onde a_1, a_2 e a_3 indicam, respectivamente, os parâmetros esquerdo, central e direito de conjunto fuzzy triangular A . De modo análogo, a mesma definição se aplica ao conjunto fuzzy triangular B .

Antes de definir os valores de cada célula na matriz de distâncias entre as respectivas amostras do segmento e do protótipo, os valores obtidos a partir da Equação 5.2 devem ser ponderados.

Considerando os conceitos do alisamento exponencial simples, foram realizados diversos experimentos com o objetivo de dar maior ênfase ao comportamento mais recente do segmento no cálculo da dissimilaridade. Entretanto, os testes demonstraram que em segmentos longos o alisamento exponencial penaliza excessivamente as primeiras amostras no cálculo da dissimilaridade pela medida DTW ponderada. Ao optarmos por atribuir maior influência ao início do segmento, as últimas amostras do segmento acabam sendo penalizadas. Outro problema do alisamento exponencial simples é a incerteza do quanto será a influência do início ou do final do segmento a partir da definição o fator de suavização indicado por α .

Diante deste contexto, a Equação 5.3 apresenta uma nova técnica de alisamento que permite a definição da influência numérica real desejada ao início do segmento.

$$w_{ij} = \log_{p0} \frac{1}{n+m} \sqrt{\frac{i+j}{n+m}} \quad (5.3)$$

onde w_{ij} é o grau de influência, ou seja, o valor utilizado na ponderação da célula ij , n é o número de linhas da matriz de distâncias entre o segmento e o protótipo, m é o

número de colunas, e $0 < p_0 < 1$ é o valor definido pelo usuário para ponderar a diferença entre as amostras na célula (1,1) da matriz, ou seja, a diferença entre os dois primeiros dados observados das sequências analisadas.

A nova técnica de alisamento considera que, respeitando a restrição $0 < p_0 < 1$, a célula (m, n) da matriz será ponderada com o valor 1. Desta forma, por meio do parâmetro p_0 , é possível definir com maior controle qual será o grau de influência das amostras iniciais do segmento e do protótipo no preenchimento da matriz de distâncias DTW.

A Figura 5.6 apresenta os graus de influência para as células da diagonal principal em uma matriz DTW com 50 linhas e 50 colunas. O valor de w_{ij} é representado pelo eixo Y da figura.

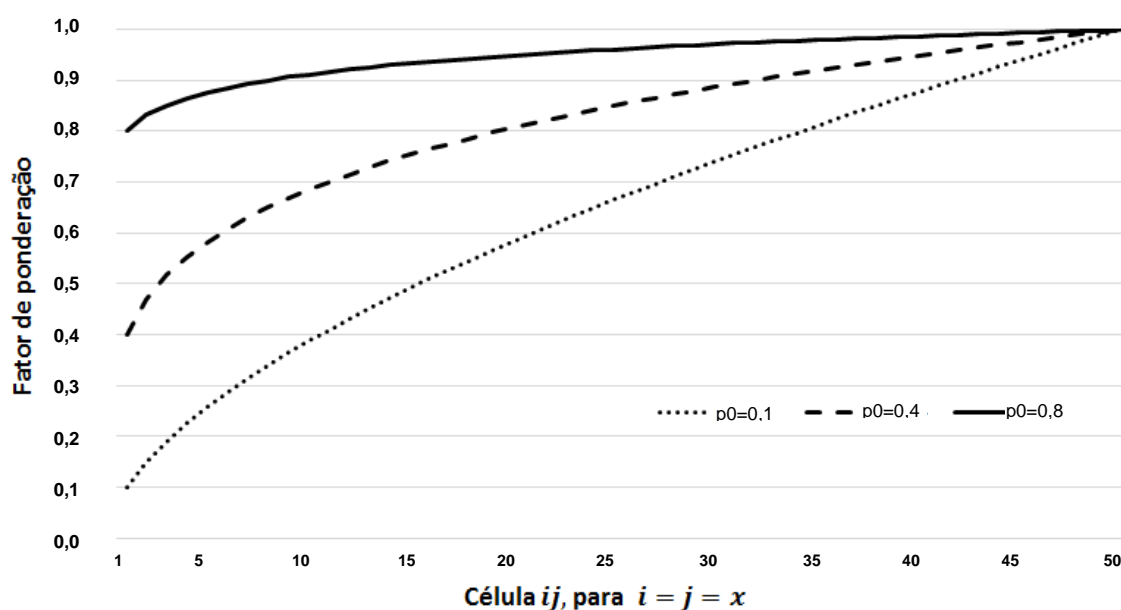


Figura 5.6 – Graus de influência das células da diagonal principal em uma matriz DTW para $p_0 = 0,1$, $p_0 = 0,4$ e $p_0 = 0,8$

A Figura 5.6 ilustra o comportamento dos valores de ponderação da diagonal principal em uma matriz 50 x 50, quando a influência para a célula (1,1) é definida com os valores 0,1, 0,4 e 0,8. Os experimentos realizados comprovam que a técnica de suavização apresentada mantém a flexibilidade na definição do caminho DTW, além de penalizar distorções indesejadas como as ilustradas na Figura 3.7. A medida que o valor definido para p_0 se aproxima de 1, mais tênue serão as diferenças dos graus de influência entre células vizinhas na matriz DTW. Para evitar que distorções ocorram

durante a análise, a restrição de Itakura (ITAKURA, 1975), apresentada na Seção 3.1.6 e ilustrada na Figura 3.8(b), também foi adotada.

5.3.2 Cálculo da pertinência do segmento de uma ST nos grupos representados pelos protótipos

O processo descrito na Seção 5.3.1 é aplicado para definir a dissimilaridade de cada segmento das ST para os intervalos correspondentes de todos os protótipos definidos como descrito na Seção 5.1.

Após o cálculo destes valores, o grau de pertinência do segmento nos grupos representados pelos protótipos é calculado como apresentado na Equação 3.14, que define pertinência de um objetivo em um respectivo grupo no algoritmo FCM. Entretanto, $x_k - c_i$ deve ser substituído pelo valor da célula (m, n) da matriz DTW definida na Seção 5.3.1, que indica a dissimilaridade entre o segmento k da série temporal analisada em relação ao intervalo correspondente do protótipo i . O maior valor $A_i(x_k)$ fornecido pela Equação 3.14, indica qual grupo melhor representa o segmento analisado.

5.4 Considerações finais

A abordagem desenvolvida para o agrupamento dinâmico das ST traz contribuições em relação a definição dos protótipos, a segmentação das ST e, em especial, a métrica utilizada para definir a dissimilaridade entre os segmentos, ou mesmo entre séries temporais.

No próximo capítulo são apresentados os experimentos realizados, os resultados obtidos e uma análise das abordagens de previsão e agrupamento desenvolvidas.

Capítulo 6

EXPERIMENTOS E AVALIAÇÃO DAS ABORDAGENS DESENVOLVIDAS

Neste capítulo são apresentados os resultados de alguns experimentos realizados para a validação do método de pré-processamento e dos modelos de previsão e de agrupamento desenvolvidos.

6.1 Experimentos para validação do pré-processamento

Os experimentos apresentados nesta seção tem como objetivo validar a primeira abordagem de pré-processamento, descrita na Seção 4.1.1. Para demonstrar a eficácia do pré-processamento desenvolvido (SANTOS; CAMARGO, 2013), foram realizados experimentos utilizando o modelo de previsão proposto por Cheng, et al. (2008) sem o pré-processamento, como na proposta original dos autores, e com o pré-processamento descrito na Seção 4.1.1.

Dentre outros experimentos realizados o modelo foi validado utilizando uma ST que indica o número de alunos matriculados anualmente na Universidade do Alabama entre os anos de 1971 e 1992. Para poder avaliar o desempenho do modelo de pré-

processamento, as previsões foram realizadas considerando os dados originais da série, sem *outliers*, e considerando a presença de um *outlier* inserido manualmente. Ambas as ST utilizadas nos experimentos, ou seja, com e sem *outlier*, são apresentadas na Tabela 6.1.

Tabela 6.1 – Séries temporais com o número anual de matrículas

Ano	Matrículas	Matrículas com <i>outlier</i>
1971	13055	13055
1972	13563	13563
1973	13867	13867
1974	14696	14696
1975	15460	15460
1976	15311	15311
1977	15603	15603
1978	15861	15861
1979	16807	16807
1980	16919	16919
1981	16388	16388
1982	15433	15433
1983	15497	15497
1984	15145	25145*
1985	15163	15163
1986	15984	15984
1987	16859	16859
1988	18150	18150
1989	18970	18970
1990	19328	19328
1991	19337	19337
1992	18876	18876

**outlier*

A partir dos dados apresentados na Tabela 6.1, foram realizadas previsões utilizando a abordagem apresentada em (CHENG, et al., 2008), com e sem o pré-processamento descrito na Seção 4.1.1. Os resultados obtidos são apresentados na Tabela 6.2. Para comparar a acurácia do modelo de previsão com e sem o pré-processamento dos dados, foram calculados o erro médio quadrático (EMQ) e a taxa média de erro (TME) . A acurácia obtida em cada experimento é apresentada na Tabela 6.3.

Tabela 6.2 – Número de matrículas anuais reais e previstas

Ano	Matrículas	Previsões sem pré-processamento		Previsões com pré-processamento	
		Com outlier	Sem outlier	Com outlier	Sem outlier
1971	13055				
1972	13563	14377	14242	13588	13715
1973	13867	14377	14242	13588	14206
1974	14696	14377	14242	14696	14206
1975	15460	15778	15474	15373	15498
1976	15311	17920	15474	15595	15566
1977	15603	15778	15474	15595	15566
1978	15861	17920	15474	16339	15566
1979	16807	16192	16146	16339	16862
1980	16919	17307	16988	17133	17133
1981	16388	17307	16988	17133	17133
1982	15433	16192	16146	15373	15498
1983	15497	17920	15474	15595	15566
1984	15145	17920	15474	*	15566
1985	15163	15135	15474	15595	15566
1986	15984	15778	15474	15595	15566
1987	16859	16192	16146	16339	16862
1988	18150	17307	16988	17133	17133
1989	18970	18971	19144	18923	18923
1990	19328	18971	19144	19332	19333
1991	19337	18971	19144	19128	19128
1992	18876	18971	19144	19128	19128

*sem valor previsto, outlier removido

Tabela 6.3 – Comparação da acurácia obtida pelas previsões com e sem o pré-processamento

Medida	Previsões sem pré-processamento		Previsões com pré-processamento	
	Com outlier	Sem outlier	Com outlier	Sem outlier
TME	5,23%	2,40%	1,70%	1,57%
EMQ	1390438	228918	148992	133513

Como é possível observar na Tabela 6.3, a representação adequada dos dados fuzzificados é capaz de melhorar a acurácia das previsões nas STF. Em situações onde há desvios nos dados observados, como ocorre nos experimentos onde um *outlier* foi inserido nas amostras, a melhora na acurácia das previsões proporcionada pelo pré-processamento torna-se mais significativa.

6.2 Experimentos com previsão

Para a validação dos modelos de previsão e análise dos resultados obtidos, foram utilizadas diversas ST com dados financeiros, econômicos, de demanda de energia, temperatura, dentre outras. Entretanto, quatro foram as séries utilizadas para a comparação da acurácia obtida com outras abordagens. A escolha das séries temporais com a cotação diária dos índices TAIEX entre os anos de 2001 e 2004 (YAHOO! FINANCE, 2013), ocorreu devido aos diferentes padrões comportamentais observados nas ST e a ampla disponibilidade de experimentos publicados na literatura com os resultados das previsões para estas séries.

Afim de obter uma análise comparativa dos resultados frente as outras abordagens disponíveis na literatura, os experimentos com os índices TAIEX foram realizados ano a ano de modo independente. Assim como ocorreu nas demais abordagens utilizadas na análise comparativa, os dados de janeiro à outubro foram utilizados para o treinamento do modelo, e os testes de previsão foram realizados com os índices dos meses de novembro e dezembro.

Os resultados obtidos pelas três abordagens de previsão foram comparados com outros sete trabalhos por meio da raiz do erro médio quadrático (RMSE). As Seções 6.2.1, 6.2.2 e 6.2.3 apresentam os dados reais e previstos para cada abordagem proposta. No final da Seção 6.2.3 é apresentada uma análise dos resultados obtidos comparados com outras abordagens disponíveis na literatura.

6.2.1 Resultados e análise dos experimentos com previsões baseadas no intervalo temporal das amostras

Os resultados apresentados nesta seção foram obtidos a partir do modelo de previsão descrito na Seção 4.2.1. O primeiro passo na tarefa de previsão é realizar o pré-processamento dos dados observados. O número de amostras, o domínio e os termos linguísticos para cada ST definidos a partir do pré-processamento apresentado na Seção 4.1.2, são apresentados na Tabela 6.4.

Tabela 6.4 – Pré-processamento nas ST dos índices TAIEX entre 2001 e 2004

Série Temporal	Amostras	Domínio	Termos Lin- guísticos	Amostras	
				Treinamento	Testes
TAIEX 2001	244	[2732,30; 6818,20]	52	201	43
TAIEX 2002	248	[3190,59; 7121,75]	49	205	43
TAIEX 2003	249	[3580,74; 6666,89]	49	206	43
TAIEX 2004	250	[4861,44; 7489,53]	48	205	45

Nos experimentos realizados, os graus de influência dos grupos de RLF foram definidos empiricamente. Os valores que indicaram os melhores resultados foram $p_{G_1} = 0,22$, $p_{G_2} = 0,33$ e $p_{G_3} = 0,44$. A Tabela 6.5 apresenta os valores reais e previstos para os índices TAIEX 2001 nos meses de novembro e dezembro.

Tabela 6.5 – Índices TAIEX 2001 reais e previstos para os meses de novembro e dezembro

Data	Valor real	Valor previsto	Data	Valor real	Valor previsto
2001/11/01	3929,69	3900,91	2001/12/03	4646,61	4453,63
2001/11/02	3998,48	3900,91	2001/12/04	4766,43	4659,84
2001/11/05	4080,51	3915,35	2001/12/05	4924,56	4780,37
2001/11/06	4082,92	3987,10	2001/12/06	5208,86	4895,68
2001/11/07	4158,15	4065,64	2001/12/07	5333,93	5227,90
2001/11/08	4135,03	4065,64	2001/12/10	5321,28	5227,90
2001/11/09	4123,78	4145,09	2001/12/11	5273,97	5309,69
2001/11/12	4172,63	4145,09	2001/12/12	5539,31	5309,69
2001/11/13	4136,54	4145,09	2001/12/13	5407,54	5588,32
2001/11/14	4277,70	4184,87	2001/12/14	5486,73	5406,05
2001/11/15	4403,59	4145,09	2001/12/17	5456,15	5515,41
2001/11/16	4446,62	4335,51	2001/12/18	5329,19	5440,46
2001/11/19	4548,63	4410,30	2001/12/19	5221,96	5309,69
2001/11/20	4455,80	4453,63	2001/12/20	5309,10	5227,90
2001/11/21	4533,37	4552,98	2001/12/21	5109,24	5227,90
2001/11/22	4450,02	4453,63	2001/12/24	5164,73	5084,22
2001/11/23	4519,08	4528,65	2001/12/25	5372,81	5175,34
2001/11/26	4608,32	4453,63	2001/12/26	5392,43	5406,05
2001/11/27	4580,33	4509,75	2001/12/27	5332,98	5380,61
2001/11/28	4447,58	4591,49	2001/12/28	5398,28	5421,21
2001/11/29	4465,83	4591,49	2001/12/31	5551,24	5372,07
2001/11/30	4441,12	4453,63	-	-	-

Os dados reais e previstos para as ST TAIEX entre os anos de 2002 e 2004 são apresentados, respectivamente, nos gráficos das Figuras 6.1, 6.2 e 6.3.

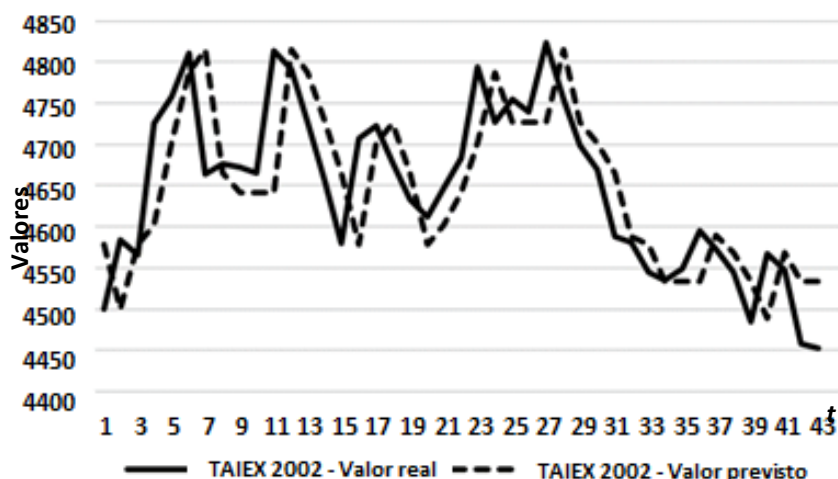


Figura 6.1 – Índices TAIEX reais e previstos para os meses de novembro e dezembro de 2002

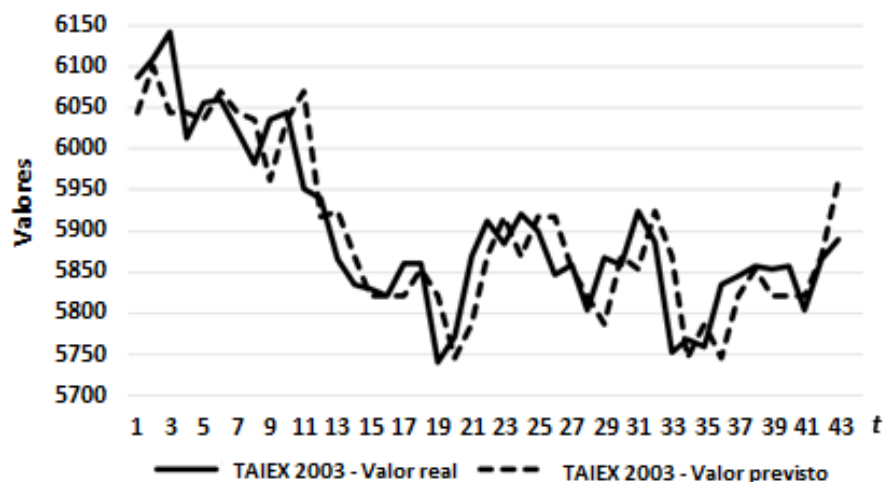


Figura 6.2 – Índices TAIEX reais e previstos para os meses de novembro e dezembro de 2003

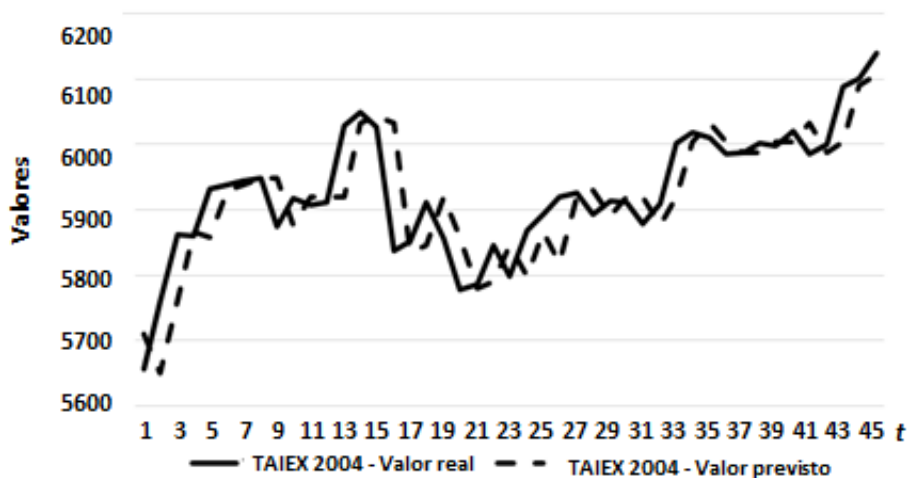


Figura 6.3 – Índices TAIEX reais e previstos para os meses de novembro e dezembro de 2004

Como é possível observar nos resultados apresentados na Tabela 6.5 e nas Figuras 6.1, 6.2 e 6.3, além de uma acurácia satisfatória, o comportamento dos valores previstos é similar ao comportamento apresentado pelos valores reais.

6.2.2 Resultados e análise dos experimentos com previsões baseadas na combinação de STF e o alisamento exponencial simples

Para validar a abordagem apresentada na Seção 4.2.2 e comprovar a evolução do modelo de previsão em relação ao modelo anterior, os experimentos para análise e publicação dos resultados foram realizados com mesmas ST utilizadas anteriormente. Para cada ano dos índices TAIEX, o treinamento do modelo foi realizado com os dados de janeiro à outubro e os testes foram realizados com os índices de novembro e dezembro.

O pré-processamento é o mesmo aplicado na abordagem de previsão da seção anterior, portanto, com as mesmas definições dos domínios e da quantidade de termos linguísticos para cada ST, conforme apresentado na Tabela 6.4.

Após a fuzzificação dos dados são extraídas as RLF de segunda ordem com três consequentes para compor a base de conhecimento. Definida a base de conhecimento inicial, as previsões são realizadas e, a cada nova amostra disponível, a base de conhecimento é atualizada com a inclusão uma nova RLF. Os valores reais e previstos para os índices TAIEX de 2003 nos meses de novembro e dezembro são apresentados na Tabela 6.6.

Nos experimentos realizados, os parâmetros α_1 e α_2 das Equações 4.9 e 4.10 foram definidos, respectivamente, com os valores 0,1 e 0,5. Os índices reais e previstos das séries TAIEX para os meses de novembro e dezembro nos anos de 2001, 2002 e 2004, são apresentados, respectivamente, nas Figuras 6.4, 6.5 e 6.6.

Os resultados obtidos nos experimentos realizados com esta abordagem de previsão, comprovam a existência de uma melhora na acurácia das previsões em relação a abordagem anterior, atribuída, principalmente, ao uso das RLF de segunda ordem com três consequentes combinadas com o alisamento exponencial simples.

Tabela 6.6 – Índices TAIEX reais e previstos para os meses de novembro e dezembro de 2003

Data	Real	Previsto	Data	Real	Previsto
2003/11/03	6087,45	6044,48	2003/12/03	5884,97	5917,75
2003/11/04	6108,99	6094,29	2003/12/04	5920,46	5869,14
2003/11/05	6142,32	6053,94	2003/12/05	5900,05	5917,75
2003/11/06	6013,40	6053,94	2003/12/08	5847,15	5917,75
2003/11/07	6056,83	6038,06	2003/12/09	5859,56	5853,37
2003/11/10	6059,03	6066,15	2003/12/10	5803,42	5799,73
2003/11/11	6022,08	6024,02	2003/12/11	5867,05	5817,84
2003/11/12	5982,75	6038,06	2003/12/12	5858,32	5869,14
2003/11/13	6035,44	5957,99	2003/12/15	5924,24	5853,37
2003/11/14	6044,77	6038,06	2003/12/16	5887,23	5925,46
2003/11/17	5952,32	6063,99	2003/12/17	5752,01	5818,87
2003/11/18	5939,47	5936,21	2003/12/18	5768,76	5750,26
2003/11/19	5865,51	5947,49	2003/12/19	5759,23	5791,74
2003/11/20	5834,24	5869,14	2003/12/22	5835,11	5791,74
2003/11/21	5830,06	5818,87	2003/12/23	5845,51	5818,87
2003/11/24	5821,58	5818,87	2003/12/24	5857,87	5833,78
2003/11/25	5861,18	5818,87	2003/12/25	5853,70	5812,87
2003/11/26	5860,61	5853,37	2003/12/26	5857,21	5812,87
2003/11/27	5740,57	5837,30	2003/12/29	5804,89	5812,87
2003/11/28	5771,77	5750,26	2003/12/30	5866,75	5873,35
2003/12/01	5870,17	5709,97	2003/12/31	5890,69	5861,44
2003/12/02	5911,45	5869,14	-	-	-

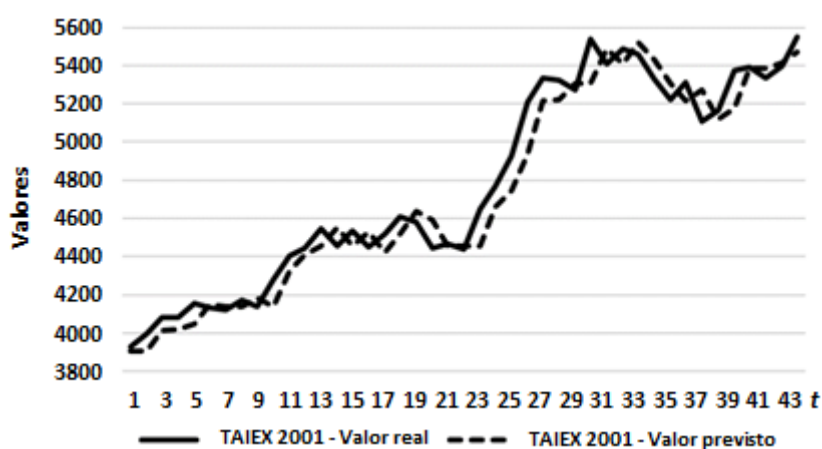


Figura 6.4 – Índices TAIEX reais e previstos para os meses de novembro e dezembro de 2001

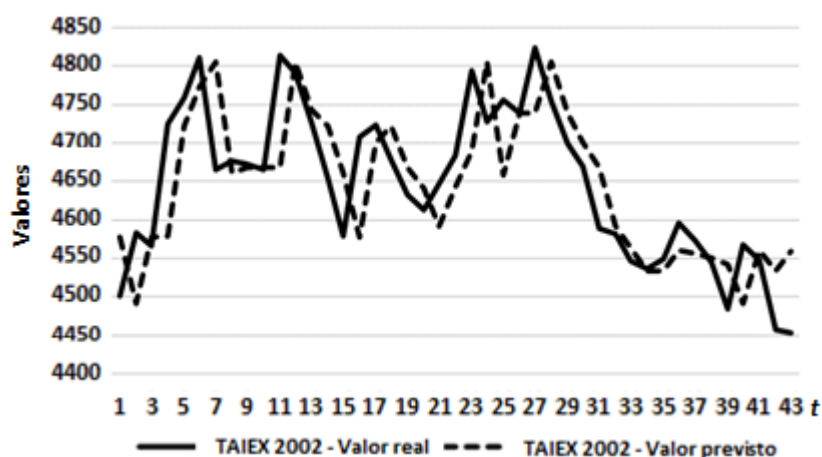


Figura 6.5 – Índices TAIEX reais e previstos para os meses de novembro e dezembro de 2002

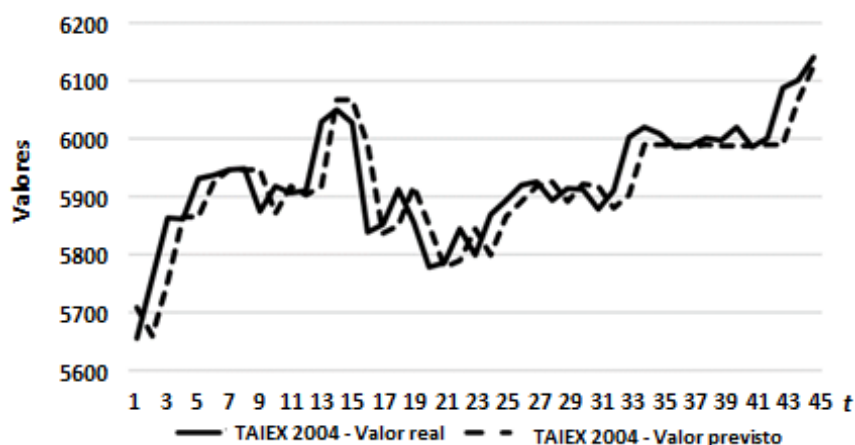


Figura 6.6 – Índices TAIEX reais e previstos para os meses de novembro e dezembro de 2004

A próxima seção apresenta os resultados obtidos pelo modelo de previsão descrito na Seção 4.2.3, além de um quadro comparativo com os índices da acurácia obtida pelos três modelos desenvolvidos comparada com outras sete abordagens da literatura.

6.2.3 Resultados e análise dos experimentos com previsões realizadas a partir da combinação de uma nova técnica de alisamento, dos conceitos de STF e da regressão linear simples

Os resultados apresentados nesta seção foram obtidos por meio do modelo de previsão descrito na Seção 4.2.3. Diferente das duas abordagens anteriores, o pré-

processamento utilizado neste modelo de previsão foi o descrito na Seção 4.1.3. Os ajustes realizados no pré-processamento, a nova técnica de suavização dos valores e os conceitos de regressão linear simples aplicados ao modelo de previsão, contribuíram diretamente para a melhora na acurácia dos resultados obtidos.

Do mesmo modo como ocorreu nas duas abordagens anteriores, os experimentos apresentados nesta seção também foram realizados utilizando os índices TAIEX entre os anos de 2001 e 2004. Para cada ST o treinamento do modelo foi realizado com os dados disponíveis entre os meses de janeiro e outubro, e os testes foram realizados com os índices de novembro e dezembro dos respectivos anos. A Tabela 6.7 apresenta os resultados obtidos pelo pré-processamento descrito na Seção 4.1.3.

Tabela 6.7 – Pré-processamento nas ST após ajustes na definição dos termos linguísticos

Série Temporal	Amostras	Domínio	Termos Linguísticos	Amostras	
				Treinamento	Testes
TAIEX 2001	244	[2732,30; 6818,20]	102	201	43
TAIEX 2002	248	[3190,59; 7121,75]	86	205	43
TAIEX 2003	249	[3580,74; 6666,89]	91	206	43
TAIEX 2004	250	[4861,44; 7489,53]	108	205	45

Considere θ_1 como o fator de suavização aplicado aos consequentes das RLF para a definição de w_i na Equação 4.13, e θ_2 como o fator de suavização para a definição de w_i na Equação 4.14. Os valores de θ_1 e θ_2 utilizados em cada um dos quatro experimentos são apresentados na Tabela 6.8.

Tabela 6.8 – Fatores de suavização utilizados nos experimentos

Série Temporal	θ_1	θ_2
TAIEX 2001	0,5	1,2
TAIEX 2002	0,4	0,2
TAIEX 2003	0,1	0,1
TAIEX 2004	0,4	7,0

Se não houver ao menos uma RLF na base de conhecimento para o cálculo do valor previsto, a previsão é realizada com o auxílio da regressão linear. Como mencionado na Seção 4.2.3, para calcular o valor de $\hat{\beta}_1$, foi considerado 20% (vinte por cento) do número total de amostras utilizadas para treinamento. A série temporal TAIEX 2002 possui 205 amostras disponíveis para treinamento, assim, em ocasiões

nas quais não havia ao menos uma RLF para o cálculo do valor previsto, foram utilizadas 41 amostras no cálculo de $\hat{\beta}_1$.

Como exemplo, para a previsão do dia 25 de novembro de 2002, não há nenhuma RLF na base de conhecimento que possa ser utilizada no cálculo. Assim, o parâmetro $\hat{\beta}_1$ foi definido considerando as 16 amostras disponíveis na base de teste, ou seja, as 16 amostras observadas entre 01 e 24 de novembro, e as 25 amostras mais recentes da base de treinamento, totalizando 41 amostras. Os valores reais e os valores previstos pelo modelo para os índices TAIEX entre 2001 e 2004 são apresentados nas Figuras 6.7, 6.8, 6.9 e 6.10, respectivamente.

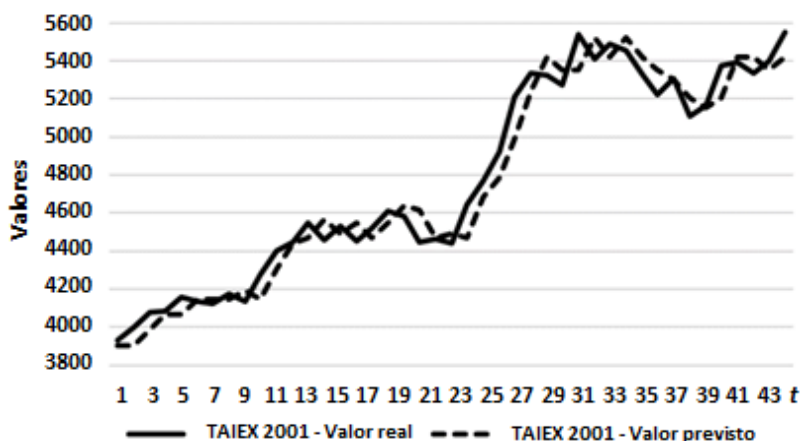


Figura 6.7 – Índices TAIEX reais e previstos para os meses de novembro e dezembro de 2001

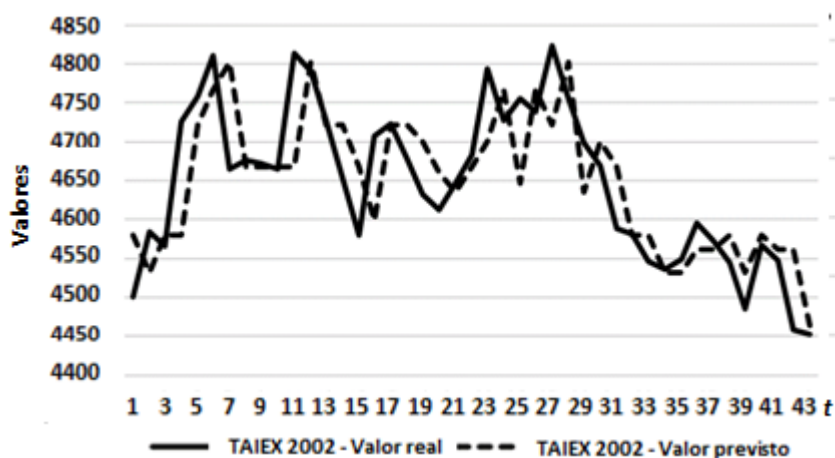


Figura 6.8 – Índices TAIEX reais e previstos para os meses de novembro e dezembro de 2002

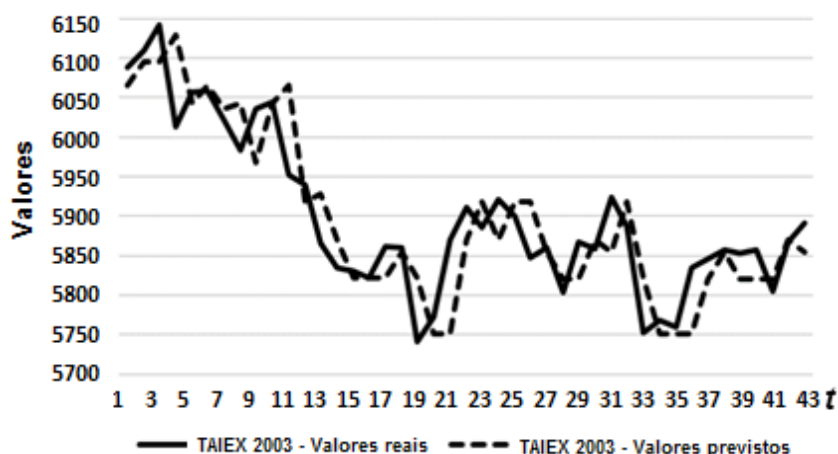


Figura 6.9 – Índices TAIEX reais e previstos para os meses de novembro e dezembro de 2003

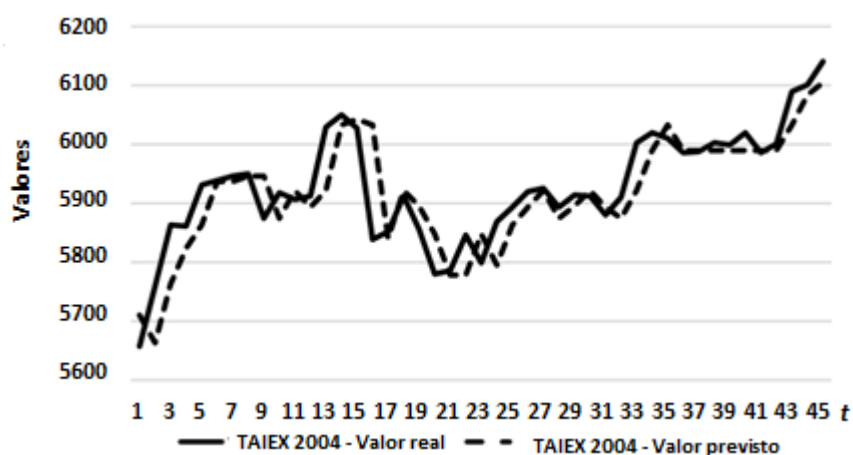


Figura 6.10 – Índices TAIEX reais e previstos para os meses de novembro e dezembro de 2004

Além dos valores previstos por este modelo de previsão também apresentarem um comportamento semelhante ao observado nos valores reais, é possível identificar uma melhora na acurácia das previsões. A Tabela 6.9 apresenta a raiz erro médio quadrático das previsões apresentadas nesta seção, com e sem o auxílio dos conceitos da regressão linear. Para permitir uma análise comparativa do desempenho desta abordagem em relação as outras duas desenvolvidas nesta tese, a Tabela 6.9 também apresenta a raiz do erro médio quadrático das previsões apresentadas nas Seções 6.2.1 e 6.2.2, além do desempenho de outras abordagens disponíveis na literatura.

A primeira abordagem da Tabela 6.9 foi desenvolvida pelos autores Brown e Meyer (1961), e apresenta os resultados obtidos pelo método do alisamento exponencial simples. A abordagem descrita em (WANG; LEE, 2014) apresenta uma proposta

de uso das redes neurais artificiais do tipo *Multilayer Perceptron* para as previsões. As demais abordagens realizam as previsões com métodos desenvolvidos a partir dos conceitos base de STF. Os resultados da abordagem (SANTOS, CAMARGO, 2014a) são referentes ao modelo de previsão apresentado na Seção 4.2.1, e a abordagem (SANTOS, CAMARGO, 2014b) traz os resultados da acurácia obtida pelo modelo de previsão descrito na Seção 4.2.2. Os resultados apresentados nas duas últimas linhas da tabela (SANTOS; CAMARGO, NO PRELO) são referentes a abordagem descrita na Seção 4.2.3, com e sem o uso de regressão linear.

Tabela 6.9 – Raiz do erro médio quadrático das previsões para os índices TAIEX com diferentes abordagens

Abordagem	TAIEX 2001	TAIEX 2002	TAIEX 2003	TAIEX 2004	Média
Brown e Meyer (1961)	123,1	66,0	52,6	56,9	74,7
Chen e Chen (2011)	115,3	71,0	58,0	57,3	75,4
Chen e Chen (2014)	114,6	64,7	53,6	52,9	71,4
Chen e Kao (2013)	114,5	76,8	54,2	58,1	75,9
Li e Chiang (2013)	115,8	64,3	57,6	55,5	73,3
Wang e Lee (2014)	120,2	69,7	56,3	54,2	75,1
Yu e Huarng (2008)	120,0	69,0	52,0	60,0	75,2
Santos e Camargo (2014a)	113,0	65,3	53,9	77,4	77,4
Santos e Camargo (2014b)	106,2	66,7	51,1	52,9	69,2
Santos e Camargo (no prelo) ¹	107,5	64,8	49,2	52,2	68,4
Santos e Camargo (no prelo) ²	97,2	63,2	48,9	52,2	65,3

¹Sem regressão linear; ²Com regressão linear

Apesar de apresentar uma acurácia compatível com os modelos de previsão disponíveis na literatura, os resultados da primeira abordagem (SANTOS, CAMARGO, 2014a) são consideravelmente inferiores aos das duas outras abordagens desenvolvidas nesta tese. A melhora na acurácia obtida em (SANTOS, CAMARGO, 2014b) é atribuída a três fatores. O primeiro deles é o uso do alisamento exponencial simples como técnica de suavização, ao invés da classificação das RLF em grupos. O segundo fator que contribuiu para a melhora da acurácia foi a representação do conhecimento por meio das RLF de segunda ordem com 3 consequentes e, por fim, a atualização constante da base de conhecimento também contribuiu significativamente para a melhora da acurácia.

Em uma análise a respeito dos dois resultados apresentados para a abordagem descrita na Seção 4.2.3 (SANTOS; CAMARGO, NO PRELO), uma característica que permitiu a melhora na acurácia dos resultados foram os ajustes realizados no pré-

processamento utilizado pelo modelo de previsão, onde conjuntos fuzzy foram adicionados nos extremos do domínio, implicando diretamente na melhora das previsões com valores abaixo de y_{min} ou acima de y_{max} . Além da adição dos novos conjuntos, a melhora também está relacionada ao uso da técnica de alisamento desenvolvida. Como é possível observar, há uma melhora na acurácia da abordagem (SANTOS; CAMARGO, NO PRELO)² apresentada na Tabela 6.9, quando comparada com os resultados da abordagem (SANTOS; CAMARGO, NO PRELO)¹. Esta melhora é atribuída ao uso da regressão linear como ferramenta de apoio nas previsões.

Entretanto, há uma situação na qual a regressão linear pode prejudicar a acurácia da previsão. Considerando que a regressão linear é utilizada quando não há RLF disponíveis na base de conhecimento para o cálculo do valor previsto e que, nestas situações, o cálculo é realizado como apresentado na Equação 5.2 considere a Figura 6.11.

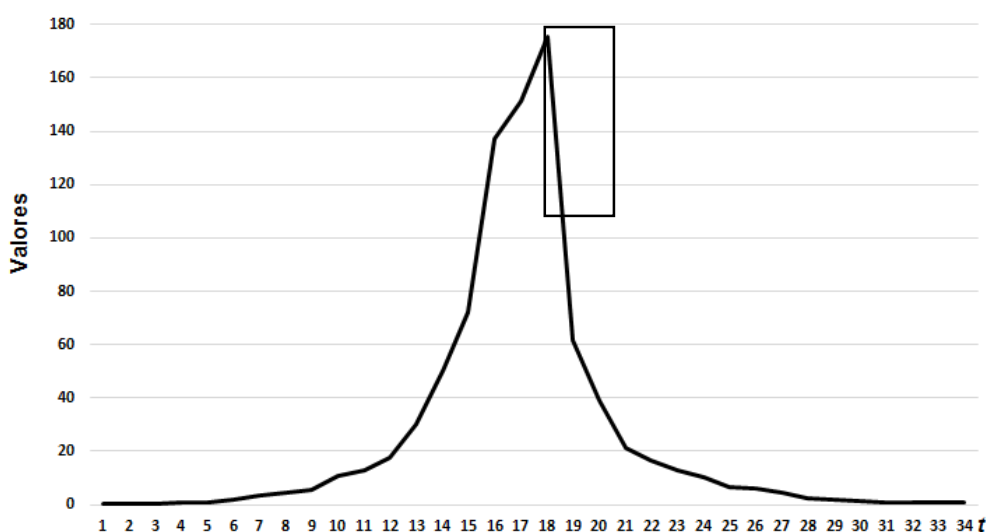


Figura 6.11 – Cenário onde ocorre um desvio pela previsão com o auxílio da regressão linear simples

Em situações como a destacada na Figura 6.11, as primeiras previsões realizadas após a mudança no comportamento da série ainda estarão considerando a tendência no crescimento dos valores, aumentando o desvio das previsões. Este problema pode ser amenizado diminuindo o número de amostras utilizadas para o cálculo de $\hat{\beta}_1$, entretanto, as previsões iniciais após a mudança comportamental sempre terão um desvio maior do que se fossem realizadas sem o auxílio da regressão linear. Por outro lado, as demais previsões na ST terão uma acurácia maior com o auxílio da

regressão linear. O problema apresentado na Figura 6.11 ocorre apenas quando não há RLF para o cálculo do valor previsto.

Outra característica dos experimentos apresentados nesta seção está relacionada aos valores de θ_1 e θ_2 na Tabela 6.8. Inicialmente, antes mesmo da primeira abordagem ser desenvolvida, acreditava-se que ao atribuir maior influência ao comportamento mais recente das ST, os resultados das previsões sempre apresentariam maior acurácia que ao atribuir maior influência ao comportamento mais antigo da série. Porém, como é possível observar na Tabela 6.8, para os índices TAIEX 2002 e 2003 a melhor acurácia foi obtida ao atribuir maior influência ao comportamento mais antigo da série.

Na próxima seção são apresentados os resultados dos experimentos realizados para a identificação de padrões comportamentais dinâmicos das séries temporais.

6.3 Experimentos com agrupamento de séries temporais

Os experimentos apresentados nesta seção foram realizados com a base de dados ECG200 (KEOGH, et al., 2011). As ST para treinamento e os protótipos definidos pelo modelo foram apresentados na Seção 5.1, respectivamente, nas Figuras 5.1 e 5.3. O domínio definido pelo pré-processamento foi $[-3,61; 5,19]$ e o número de termos linguísticos para a representação dos valores *crisp* é 47. A lista dos termos linguísticos e seus respectivos parâmetros podem ser consultados no Apêndice A.

O conjunto de dados para treinamento está dividido em dois grupos, sendo 69 ST pertencentes ao Grupo 1 e 31 ST pertencentes ao Grupo 2, totalizando 100 ST com 96 amostras cada uma. O conjunto de ST para testes também é composto por 100 séries com 96 amostras cada uma, sendo 64 do Grupo 1 e 36 do Grupo 2.

Após a fuzzificação das séries de treinamento e a definição dos protótipos, o próximo passo é fuzzificar as ST da base de testes e os protótipos, considerando os conjuntos definidos durante a fase de treinamento.

Na sequência deve ser realizada a segmentação das ST da base de testes. Quatro séries foram selecionadas para ilustrar o processo de agrupamento, sendo 2 pertencentes ao Grupo 1 e 2 pertencentes ao Grupo 2. A Tabela 6.10 apresenta o

número de segmentos definido para cada uma das quatro séries temporais considerando o *threshold* de R^2 com os valores 0,85, 0,90 e 0,95 para a segmentação.

Tabela 6.10 – Segmentos em cada ST

Série Temporal	R^2	Segmentos	Máx.de Amostras em um Segmento	Mín. de Amostras em um Segmento
ST1 – Grupo 1	0,85	20	30	2
ST2 – Grupo 1	0,85	10	26	2
ST3 – Grupo 2	0,85	16	32	2
ST4 – Grupo 2	0,85	9	28	2
ST1 – Grupo 1	0,90	24	24	2
ST2 – Grupo 1	0,90	20	24	2
ST3 – Grupo 2	0,90	19	28	2
ST4 – Grupo 2	0,90	10	28	2
ST1 – Grupo 1	0,95	26	24	2
ST2 – Grupo 1	0,95	31	22	2
ST3 – Grupo 2	0,95	25	22	2
ST4 – Grupo 2	0,95	25	18	2

Nos experimentos realizados e apresentados nesta seção, foi utilizado o *threshold* de $R^2 = 0,9$ para a segmentação das ST presentes na base de testes. A Figura 6.12 apresenta as 4 séries selecionadas e os segmentos resultantes após a execução do processo de segmentação da Seção 5.2.

Concluída a segmentação das séries temporais da base de testes, deve ser calculada a distância de cada segmento aos respectivos intervalos dos protótipos como apresentado na Seção 5.3.1. Após o cálculo da distância entre o segmento e o intervalo correspondente de todos protótipos, a pertinência do segmento nos respectivos grupos é calculada como apresentado na Seção 5.3.2. A Figura 6.13 apresenta o comportamento da ST1 da Figura 6.12, em relação aos dois protótipos que representam o comportamento das ST do Grupo 1 e do Grupo 2.

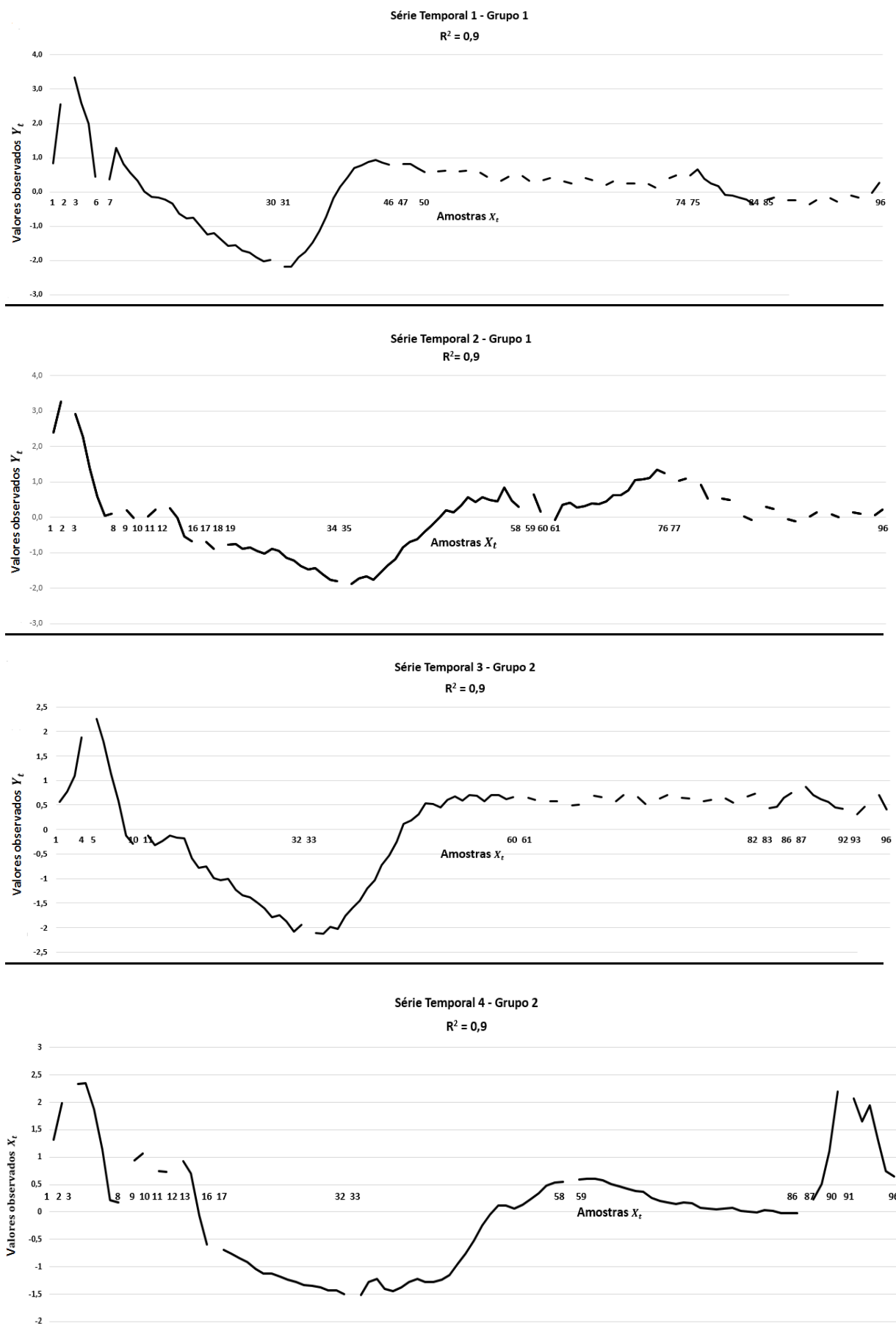


Figura 6.12 – Quatro ST da base ECG200 segmentadas

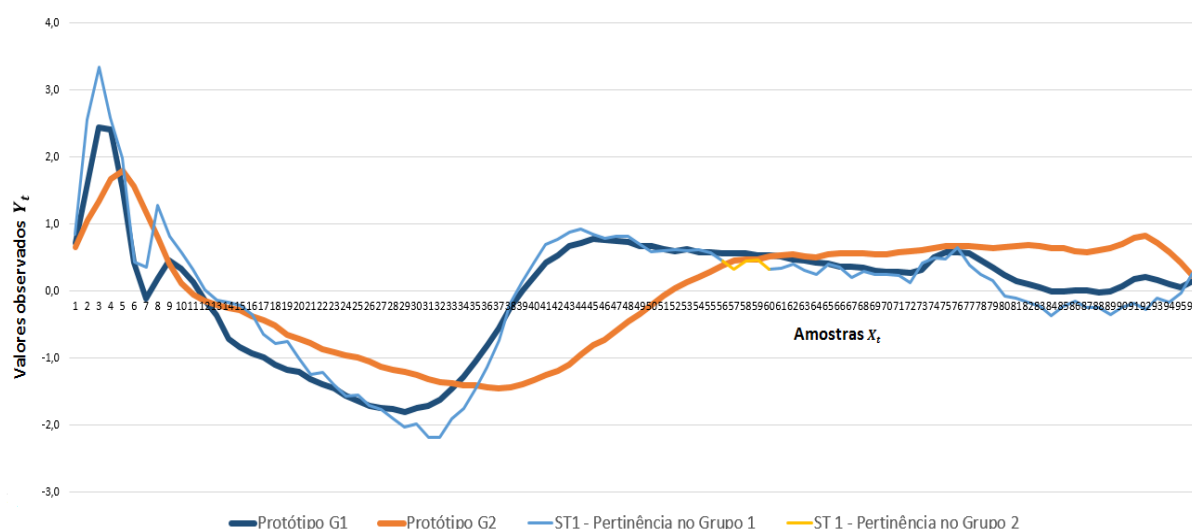


Figura 6.13 – Comportamento da ST1 em relação aos protótipos do Grupo 1 e do Grupo 2

Como é possível observar na Figura 6.13, praticamente durante todo o período de observação da ST1 ela possui o maior grau de pertinência associado ao Grupo 1, com exceção aos segmentos 9 e 10, referente ao intervalo das amostras 57 à 60 da série. A Figura 6.14 apresenta a matriz DTW que define a distância fuzzy ponderada do segmento 2 da ST1, correspondente ao intervalo das amostras 3 à 6, para o intervalo correspondente dos protótipos que representam o Grupo 1 e o Grupo 2 das ST.

		Termos linguísticos do Protótipo do G1			
		L_{46}	L_{46}	L_{43}	L_{17}
Termos linguísticos do segmento 2 da ST1	L_{46}	0,000	0,000	3,173	4,821
	L_{46}	0,000	0,000	3,323	5,007
	L_{45}	2,558	2,679	0,949	2,635
	L_{19}	4,799	4,984	1,664	0,025

(a)

		Termos linguísticos do Protótipo do G2			
		L_{43}	L_{44}	L_{44}	L_{44}
Termos linguísticos do segmento 2 da ST1	L_{46}	2,747	2,743	2,911	3,049
	L_{46}	2,989	2,911	3,049	3,167
	L_{45}	0,872	0,552	0,574	0,592
	L_{19}	1,551	2,017	2,083	2,141

(b)

Figura 6.14 – Distância ponderada entre os conjuntos fuzzy que representam os dados observados entre as amostras 3 e 6 da ST1 e os protótipos

As células (i, j) , para $i = 1, 2, 3$ e 4 ; e $j = 1, 2, 3$ e 4 das matrizes na Figura 6.14, representam a distância do i – ésimo termo fuzzy da ST1 ao j – ésimo termo fuzzy dos respectivos protótipos. Os termos linguísticos do segmento 2 na série temporal 1, no intervalo de 3 à 6 são, respectivamente, L_{46}, L_{46}, L_{45} e L_{19} . Os respectivos termos linguísticos do protótipo do Grupo 1 são L_{46}, L_{46}, L_{43} e L_{17} e do Grupo 2 L_{43}, L_{44}, L_{44} e

L_{44} . As distâncias entre os conjuntos apresentadas nas células das matrizes na Figura 6.14, foram ponderadas pelo valor w_{ij} definido pela Equação 5.3 considerando $p_0 = 0,75$.

Assim, ao preencher a matriz DTW como descrito a partir da Figura 3.4 e utilizando w_{ij} para ponderar as distâncias entre os conjuntos fuzzy da matriz, a distância entre o segmento 2 da ST1 para o protótipo do Grupo 1 é 0,97 e para o protótipo do Grupo 2 é 8,37. Ambos os valores foram definidos pela soma dos valores presentes na diagonal principal das matrizes. A Tabela 6.11 apresenta os graus de pertinência dos 24 segmentos da ST1 nos intervalos equivalentes dos protótipos do Grupo 1 e do Grupo 2.

Tabela 6.11 – Pertinência dos segmentos da ST1 nos protótipos do Grupo 1 e do Grupo 2

Segmento	Pertinência G1	Pertinência G2	Segmento	Pertinência G1	Pertinência G2
1	0,56	0,44	13	0,87	0,13
2	0,89	0,11	14	0,74	0,26
3	0,56	0,44	15	0,86	0,14
4	0,79	0,21	16	0,82	0,18
5	0,94	0,06	17	0,78	0,22
6	0,98	0,02	18	0,71	0,29
7	0,92	0,08	19	0,76	0,24
8	0,69	0,31	20	0,71	0,29
9	0,28	0,72	21	0,72	0,28
10	0,40	0,60	22	0,68	0,32
11	0,58	0,42	23	0,77	0,23
12	0,58	0,42	24	0,64	0,36

Os segmentos listados na Tabela 6.11 foram apresentados na Figura 6.12, e a representação gráfica da maior pertinência dos segmentos é ilustrada na Figura 6.13. De modo análogo, o comportamento de ST2 em relação aos dois protótipos que representam os Grupos 1 e 2, é apresentado na Figura 6.15.

Embora a ST2 seja rotulada como pertencente ao Grupo 1 representado pela linha azul escura mais espessa, é possível observar que durante um longo período de observação o comportamento da ST2 possui maior semelhança com o protótipo que representa as séries do Grupo 2.

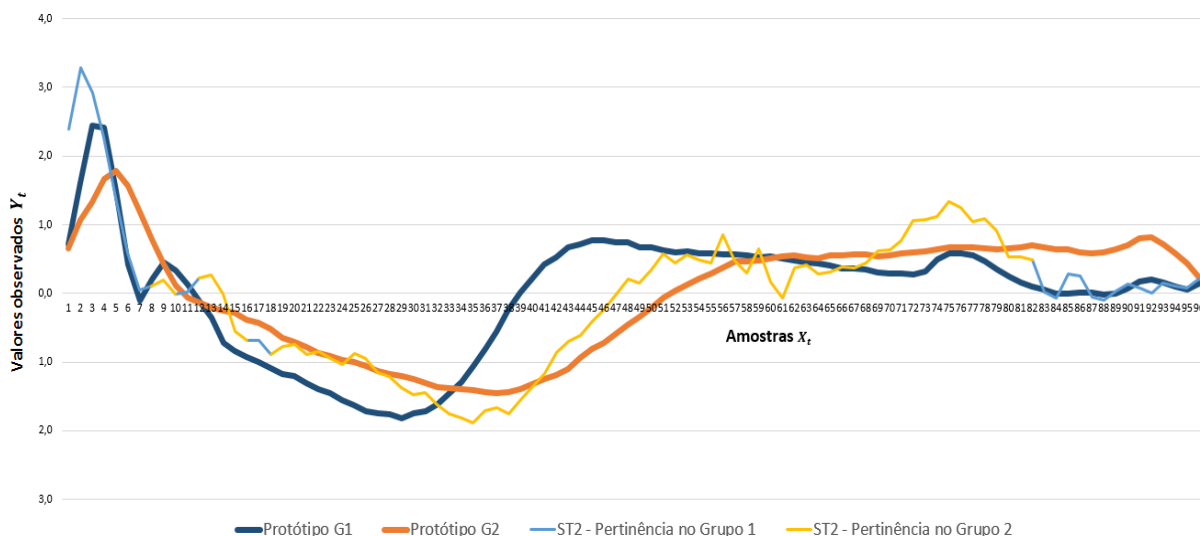


Figura 6.15 – Comportamento da ST2 em relação aos protótipos do Grupo 1 e do Grupo 2

Na Figura 6.16 é apresentada a ST3, a qual possui pertinência maior que 0,5 ao protótipo do Grupo 2 durante todo o período de observação.

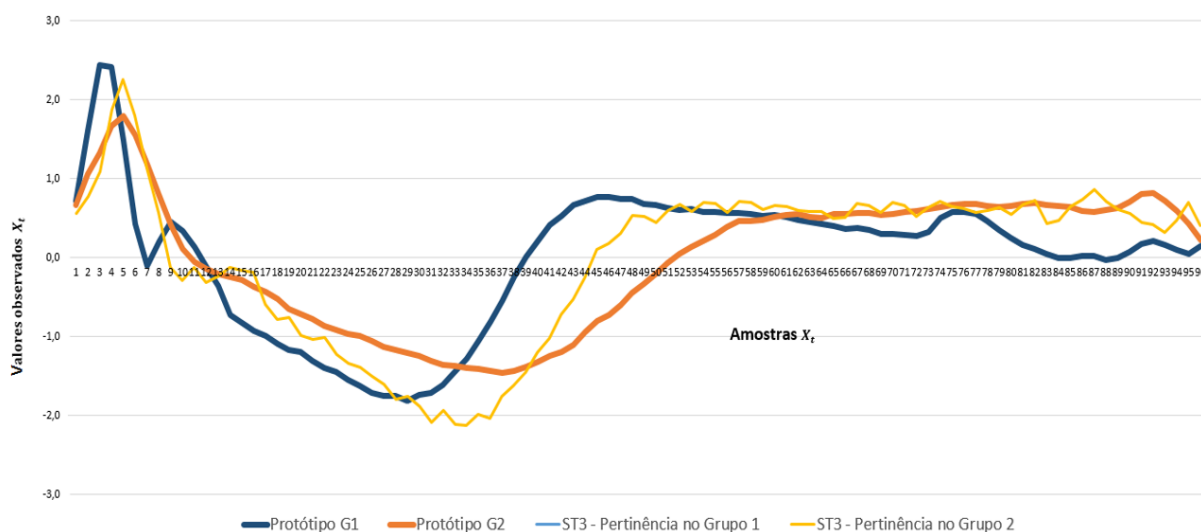


Figura 6.16 – Comportamento da ST3 em relação aos protótipos do Grupo 1 e do Grupo 2

O exemplo apresentado da Figura 6.16 ilustra o comportamento de uma ST associada, durante todo o período de observação, ao Grupo 2. Entretanto, cabe ressaltar que, embora o segmento 4, definido entre as amostras 33 e 60, tenha sido associado ao Grupo 2, o grau de pertinência do segmento ao Grupo 1 é de 0,48, o que indica uma similaridade considerável do segmento com o grupo, principalmente em razão do comportamento observado nas últimas amostras do segmento. A Figura 6.17 apresenta o comportamento da ST4 em relação protótipos do Grupo 1 e do Grupo 2.

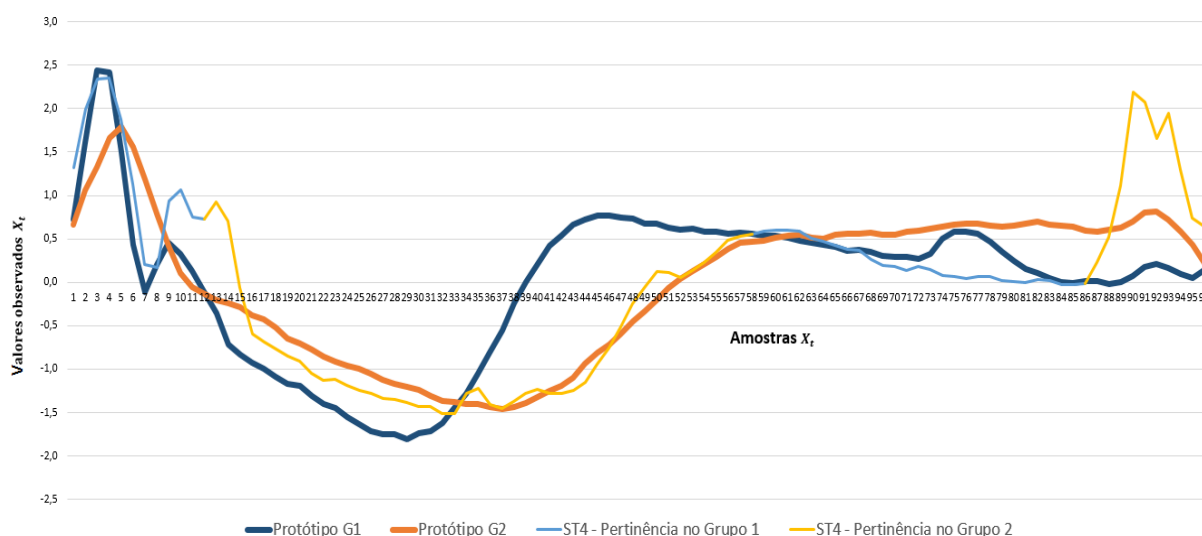


Figura 6.17 – Comportamento da ST4 em relação aos protótipos do Grupo 1 e do Grupo 2

A ST4, originalmente rotulada como pertencente ao Grupo 2, apresenta intervalos de pertinência em ambos os grupos. O comportamento no início da série é típico das ST do Grupo 1. Na sequência, a série intercala diferentes comportamentos até o final do período de observação.

A Tabela 6.12 apresenta uma análise dos resultados para as 4 séries temporais utilizadas neste capítulo para ilustrar as etapas do processo de identificação de alterações comportamentais. A análise considera o comportamento da ST de acordo com os segmentos que foram definidos e também considerando o número de amostras disponíveis.

Tabela 6.12 – Comportamento das ST apresentadas em relação aos grupos

Série Temporal	Total de Segmentos	Segmentos G1	Segmentos G2	Amostras G1	Amostras G2
ST1 – Grupo 1	24	22	2	92	4
ST2 – Grupo 1	20	12	8	28	68
ST3 – Grupo 2	19	0	19	0	96
ST4 – Grupo 2	10	5	5	40	56

A Tabela 6.12 apresenta o número de segmentos e de amostras pertencentes a cada grupo ao longo de todo o período de observação das ST. É importante ressaltar que os resultados ilustram a análise realizada nos experimentos apresentados, podendo haver variações nos números de acordo com os parâmetros R^2 e p_0 definidos.

A partir dos experimentos realizados é possível concluir que as séries podem mudar o comportamento ao longo do período de observação, o que torna a análise comportamental dinâmica uma boa alternativa para a avaliação do comportamento de uma ST. A abordagem desenvolvida nesta tese aplica a segmentação das ST com base no valor de R^2 . Assim, a definição do tamanho dos segmentos tem relação direta com o valor escolhido como *threshold*. Segmentos muito pequenos, com duas amostras por exemplo, normalmente não são suficientes para identificar uma tendência nos valores. Entretanto, como é possível observar nos exemplos, a definição de segmentos menores não comprometeu a análise comportamental realizada pela abordagem desenvolvida.

Inicialmente, a proposta de suavização apresentada para a abordagem de agrupamento tinha como objetivo avaliar os segmentos concatenados, ou seja, ao analisar o segmento 3, por exemplo, também seriam considerados na análise as amostras dos segmentos 1 e 2. Entretanto, os resultados dos testes realizados demonstraram melhor eficácia com a análise isolada, além de reduzir consideravelmente o tempo de processamento. A opção por continuar utilizando a suavização, deve-se principalmente a flexibilidade da técnica desenvolvida pois, se desejado, é possível atribuir $p_0 = 0,999$ e praticamente anular o efeito da ponderação.

6.4 Considerações finais

Os experimentos apresentados neste capítulo demonstram a importância do pré-processamento tanto na previsão de valores, quanto no agrupamento das séries temporais. A flexibilidade fornecida pela técnica de suavização do modelo de previsão, que permite atribuir maior influência ao período inicial ou final da série, também exerce influência direta na acurácia das previsões. Em relação ao modelo de agrupamento, um fator importante é o *threshold* definido para a segmentação das séries temporais. Porém, mesmo quando a análise foi realizada em pequenos segmentos, o modelo apresentou um desempenho satisfatório em relação aos resultados obtidos.

No próximo capítulo são apresentadas as conclusões, as limitações dos modelos desenvolvidos e os trabalhos futuros.

Capítulo 7

CONCLUSÕES E TRABALHOS FUTUROS

A combinação dos sistemas fuzzy com os conceitos clássicos das séries temporais tem demonstrado, não só por meio deste trabalho, ser uma área de pesquisa bastante promissora. Para chegar às abordagens apresentadas nesta tese, inúmeros experimentos foram realizados com a combinação de diferentes conceitos e, em sua maioria, com resultados compatíveis, em alguns casos melhores, que os disponíveis na literatura.

Uma característica evidenciada ao longo da pesquisa é a importância do pré-processamento. Em um mesmo modelo de previsão é possível obter previsões satisfatórias ou não, dependendo do pré-processamento aplicado. Em razão do modelo de previsão utilizar o valor do parâmetro central do conjunto triangular fuzzy como valor previsto, uma definição equivocada dos conjuntos pode produzir desvios consideráveis nas previsões. Outro ponto importante do pré-processamento é a identificação dos *outliers*. Existe um limiar muito tênue entre a identificação de um *outlier* e a presença de um novo dado que pode indicar uma mudança significativa na série. Durante os experimentos de previsão realizados para o artigo submetido para a sessão especial sobre *Computational Intelligence in Forecasting (CIF)* da *International Fuzzy*

Systems Association (IFSA), em algumas previsões a acurácia foi maior sem a remoção dos dados considerados *outliers*. Atualmente há trabalhos específicos para distinguir um *outlier* de uma novidade (DINGA, LIA, 2014), (OLIVEIRA, MEIRA, 2006).

Os modelos de previsão desenvolvidos tem como principal característica a atribuição de diferentes graus de influência aos dados observados ao longo da série. Além de considerar o número de ocorrências de cada RLF, as abordagens também consideram o momento em que cada RLF ocorre. Com o desenvolvimento do primeiro modelo de previsão, no qual foi obtida a menor acurácia dentre os três modelos desenvolvidos, os resultados sinalizaram ser uma alternativa viável a valorização do conhecimento mais recente da série temporal. Entretanto, com a evolução da abordagem para previsão e com os experimentos realizados em diversos domínios, constatou-se que nem sempre atribuir a maior influência ao conhecimento recente da série implicará em uma melhor acurácia nas previsões.

A abordagem desenvolvida para o agrupamento das séries temporais combina diferentes linhas de pesquisa. Na literatura há trabalhos voltados para a definição de protótipos em séries temporais (HAUTAMAKI, et al., 2008), trabalhos com foco na segmentação (KEOGH, et al., 2004) e trabalhos com foco no agrupamento (D'URSO, et al., 2013). Para o agrupamento dinâmico foi necessário combinar, não só estas etapas, mas também o pré-processamento, essencial para uma melhor representatividade fuzzy em base de dados como a utilizada nos experimentos desta tese.

O desenvolvimento das pesquisas abordando o pré-processamento, a previsão de valores e o agrupamento das séries temporais, proporcionaram ao trabalho contribuições importantes, além de diversas oportunidades de aprimoramento e continuidade da pesquisa. Na sequência, são apresentadas as principais contribuições, as publicações obtidas, as limitações das abordagens desenvolvidas e as oportunidades de trabalhos futuros.

7.1 Contribuições

Na etapa do pré-processamento a principal contribuição está relacionada a definição dos termos linguísticos e suas respectivas funções de pertinência. Os experimentos realizados ao longo da pesquisa utilizando o pré-processamento bem como o estudo de sua importância, comprovaram a melhora na acurácia dos valores previstos e também uma melhor representatividade dos dados *crisp* por meio dos termos linguísticos definidos.

Os modelos de previsão desenvolvidos nesta tese trazem para as séries temporais fuzzy o cálculo ponderado dos valores previstos por meio de uma nova técnica de suavização descrita na Seção 4.2.3. Além de obter resultados satisfatórios, a técnica de suavização mostrou-se coerente e de fácil interpretação em relação ao parâmetro do índice de suavização. Quando desejado, a técnica de suavização desenvolvida permite atribuir maior influência na previsão ao comportamento mais recente ou mais antigo da série. A construção das bases de conhecimento utilizando RLF com três consequentes é uma característica introduzida às STF pela abordagem descrita na Seção 4.2.2.

Os conceitos de regressão linear simples possuem fundamental importância nos modelos de previsão e agrupamento desenvolvidos. A identificação da tendência sinalizada pelas amostras mais recentes da série indicada pelo parâmetro $\hat{\beta}_1$, contribuiu significativamente para a melhora da acurácia nas previsões.

A característica do agrupamento dinâmico das ST presente na abordagem desenvolvida e a métrica elaborada para cálculo da distância entre dois segmentos, ou duas séries temporais, são as principais contribuições do modelo de agrupamento. Além de considerar os conjuntos fuzzy no cálculo da distância DTW, o que naturalmente ameniza as pequenas variações presentes nos dados observados, o algoritmo desenvolvido ainda permite, se desejado, o cálculo ponderado das distâncias. Utilizando a métrica desenvolvida associada ao algoritmo FCM, os experimentos apresentaram resultados satisfatórios em relação a análise de pertinência de um segmento nos respectivos protótipos de representação dos grupos. Como mencionado na Seção 5.2, o número e o tamanho dos segmentos estão diretamente relacionados a definição

do valor de R^2 . Embora a abordagem de segmentação desenvolvida resulte, em alguns casos, na existência de segmentos pequenos, esta característica não comprometeu a análise da pertinência destes segmentos nos respectivos protótipos.

7.2 Publicações

No início das pesquisas para a definição da estrutura base deste trabalho, foram identificadas três frentes de desenvolvimento: (1) pré-processamento; (2) previsão e (3) análise do comportamento das ST. A proposta inicial foi que, ao ser concluída cada uma das etapas, ao menos um artigo deveria ser submetido para publicação. Além das publicações indicarem de alguma forma a qualidade do trabalho desenvolvido e o quão significativo é o tema abordado, os comentários e sugestões dos avaliadores foram de fundamental importância para o aprimoramento dos modelos apresentados.

Os resultados obtidos pela abordagem de pré-processamento apresentada na Seção 4.1.1, foram publicados no artigo “*Preprocessing in Fuzzy Time Series to Improve the Forecasting Accuracy*”, apresentado na *International Conference on Machine Learning and Applications* de 2013 (ICMLA13) (SANTOS; CAMARGO, 2013), realizada em Miami. No artigo é possível concluir que a representação fuzzy adequada dos dados *crisp* influencia, de forma positiva, a acurácia das previsões realizadas.

Em relação às abordagens de previsão desenvolvidas, foram realizadas três publicações. A primeira delas, referente a abordagem de previsão apresentada na Seção 4.2.1, foi publicada no artigo intitulado “*Séries Temporais Fuzzy: Um Modelo de Previsão Baseado no Intervalo Temporal das Amostras*”, apresentado no Congresso Brasileiro de Sistemas Fuzzy (CBSF) de 2014, realizado em João Pessoa (SANTOS; CAMARGO, 2014b). A segunda abordagem, descrita na Seção 4.2.2, foi publicada e apresentada na *Ibero-American Conference on Artificial Intelligence* (IBERAMIA) de 2014, realizada em Santiago no Chile, no artigo intitulado “*Forecasting in Fuzzy Time Series by an Extension of Simple Exponential Smoothing*” (SANTOS; CAMARGO, 2014a).

A terceira e principal abordagem de previsão desenvolvida, descrita na Seção 4.2.3, foi submetida ao *16th World Congress of the International Fuzzy Systems Association (IFSA)*, no artigo intitulado “*A hybrid forecast model combining fuzzy time series, linear regression and a new smoothing technique*” (SANTOS; CAMARGO, NO PRELO). O aceite do artigo para a publicação foi recebido dias antes da entrega desta tese. Além da publicação, no IFSA também haverá uma competição entre os pesquisadores da área de séries temporais denominada *Computational Intelligence in Forecasting (CIF)*. Esta competição torna ainda mais interessante a participação no congresso, não só em função da avaliação do modelo desenvolvido, mas também por reunir uma quantidade significativa de pesquisadores da mesma área. O objetivo da competição é verificar qual método apresenta a melhor acurácia nas previsões. Para isso, foi disponibilizada uma base de dados com 91 séries temporais dos mais diversos tamanhos e comportamentos. A partir do modelo apresentado na Seção 4.2.3, e também descrito no artigo enviado ao IFSA, foram realizadas as previsões para as 91 séries, uma a uma, conforme as regras definidas pelos organizadores. Os resultados obtidos foram submetidos ao comitê avaliador e os vencedores serão divulgados durante o congresso, a ser realizado nas Astúrias na Espanha entre os dias 30 de junho e 3 de julho de 2015.

Além dos trabalhos já publicados, atualmente está sendo produzido um artigo com foco na análise comportamental das séries temporais, que apresentará a abordagem descrita no Capítulo 5 desta tese. Ao ser concluído, o artigo será submetido ao *journal Neurocomputing*.

7.3 Limitações e trabalhos futuros

A partir dos 91 experimentos realizados para o CIF, foi identificada uma limitação do modelo de previsão ligada ao tamanho das ST, mais especificamente, a quantidade de observações disponíveis para treinamento. Algumas séries disponibilizadas pelos organizadores possuem cerca de 10 à 15 amostras sendo, normalmente, de 3 à 5 para testes. Em casos como este, o modelo de previsão pode não obter uma acurácia satisfatória nos resultados. Além de possuir poucas amostras para o treinamento do modelo de previsão, o comportamento dos dados de teste em algumas ST

era completamente oposto ao comportamento dos dados para treinamento. Esta característica resultou no uso da regressão linear como ferramenta de auxílio na previsão. Entretanto, em alguns casos os dados de treinamento indicavam uma tendência de crescimento nos valores e, a partir da primeira amostra disponível para teste, os valores apresentavam tendência de diminuição, o que resultou em um aumento no desvio de algumas previsões realizadas.

Um ajuste no pré-processamento que pode amenizar o erro de previsões em situações semelhantes a descrita anteriormente, é a introdução dos conceitos relacionados a identificação de novidades. Com a abordagem de pré-processamento atual, não há uma análise dos dados para a identificação de novidades. Ao ser considerado um *outlier*, o dado simplesmente é removido.

Em relação à acurácia das previsões, há uma forte relação entre a definição dos valores de α_1 , α_2 , θ_1 e θ_2 da Tabela 6.8, e os resultados obtidos. Com este enfoque, alguns experimentos iniciais já foram realizados para o ajuste automático dos melhores valores de alfa. A proposta é identificar a partir dos dados disponíveis, quais valores de α ou θ implicariam nas melhores previsões, respectivamente, para a segunda e a terceira abordagens de previsão apresentadas. Para isso, após a chegada de cada nova amostra da base de testes deve ser identificado quais valores de alfa gerariam o menor desvio. A partir da análise histórica dos melhores alfas para as previsões já realizadas, seriam então definidos os valores de α e θ para cada nova previsão. A expectativa é que, por meio do ajuste automático dos valores de alfa, seja possível melhorar a acurácia do modelo.

Em situações onde não há ao menos uma RLF para o cálculo do valor previsto, a acurácia está relacionada ao número de amostras utilizadas para no cálculo de $\hat{\beta}_1$. Uma proposta de trabalho futuro com foco na definição do número de amostras a ser utilizada no cálculo, é realizar a segmentação da ST a partir da última amostra disponível, ou seja, identificar o último segmento linear da série e considerar o número de amostras deste segmento no cálculo de $\hat{\beta}_1$. Assim, somente a tendência apresentada pelo segmento atual será considerada na previsão.

Outro trabalho futuro relacionado a abordagem de previsão, é o uso da técnica de alisamento utilizada no agrupamento das séries temporais. Por se tratar de uma técnica de suavização onde há um controle baseado em números reais, e não na

definição de um parâmetro alfa, a expectativa é que a técnica produza previsões compatíveis, ou melhores, as já apresentadas pelos modelos desenvolvidos.

O modelo de segmentação combinado com a abordagem de agrupamento tem produzido resultados satisfatórios. Uma possibilidade para desenvolvimento de trabalho futuro é associar à análise comportamental apresentada nesta tese, a previsão de mudanças no comportamento das mesmas. O objetivo desta nova abordagem é prever o momento em que uma ST pode deixar de ter seu comportamento vinculado a um grupo, e passar a apresentar característica de outro grupo. Uma alternativa para realizar este tipo de análise é a possibilidade de que seguidas previsões de valores possam indicar, ou não, a mudança de comportamento da série. Entretanto, esta abordagem é fortemente dependente da acurácia das previsões a médio ou longo prazo. Outra abordagem possível é a análise histórica do comportamento das séries por meio das pertinências dos segmentos. Nesta abordagem, a proposta é identificar qual foi o comportamento das ST que migraram de um grupo para outro. Identificado o padrão comportamental que precedeu tais mudanças, o modelo identifica as séries com comportamentos semelhantes e sinaliza a possibilidade de migração da série entre os grupos.

Por se tratar de uma área com um grande potencial de pesquisas, inúmeras são as possibilidades de trabalhos futuros além das apresentadas nesta seção. Durante o desenvolvimento de novas abordagens, ou mesmo durante o aprimoramento das abordagens apresentadas nesta tese, certamente novos desafios surgirão. Diferentes cenários exigem diferentes ajustes dos modelos, o que muitas vezes pode resultar no desenvolvimento de novas abordagens.

REFERÊNCIAS

AGRAWAL, R., FALOUTSOS, C. & SWAMI, A., 1993. Efficient similarity search in sequence databases. *Lecture notes in computer science*, Volume 730, p. 69–84.

AGRAWAL, R., LIN, K., SAWHNEY, H. S. & SHIM, K., 1995. Fast Similarity Search in the Presence of Noise, Scaling, and Translation in Time-Series Databases. *Proceedings of the 21th International Conference on Very Large Data Bases*, pp. 490-501.

BAGNALL, A. & JANACEK, G., 2005. Clustering time series with clipped data. *Machine Learning*, Volume 58, p. 151–178.

BAGNALL, A.; RATANAMAHATANA, C.; KEOGH, E.; LONARDI, S. & JANACEK, G., 2006. A bit level representation for time series data mining with shape based similarity. *Data Mining and Knowledge Discovery*, Volume 13, p. 11–40.

BARNETT, V. & LEWIS, T., 1994. *Outliers in statistical data*. 3rd ed. NY: John Wiley & Sons.

BELLMAN, R. E., 1957. *Dynamic Programming*. s.l.:Princeton University Press.

BEZDEK, J. C., 1981. *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum Press.

BEZDEK, J. C., TSAO, E. C. & PAL, N. R., 1992. Fuzzy kohonen clustering networks. *IEEE International Conference on Fuzzy Systems*, pp. 1035-1043.

BOX, G. E. P. & JENKINS, G., 1976. *Time Series Analysis: Forecasting and Control*. Oakland: Holden-Day.

BROWN, R. & MEYER, R., 1961. The fundamental theory of exponential smoothing. *Operations Research* 9, pp. 673-685.

BUENO, R. L. S., 2008. *Econometria de Séries Temporais*. São Paulo: Cengage Learning.

CAI, Q., ZHANG, D., ZHENG, W. & LEUNG, S. C. H., 2015. A new fuzzy time series forecasting model combined with ant colony optimization and auto-regression. *Knowledge-Based Systems*, Volume 74, p. 61–68.

CHATTERJEE, S., NIGAM, S., SINGH, J. B. & UPADHYAYA, L. N., 2011. Application of Fuzzy Time Series in Prediction of Time Between Failures & Faults in

Software Reliability Assessment. *Fuzzy Information and Engineering*, Volume 3, pp. 293-309.

CHENG, C., CHENG, G. & WANG, J., 2008. Multi-attribute fuzzy time series method based on fuzzy clustering. *Expert Systems with Applications*, Volume 34, p. 1235–1242.

CHEN, S. M. & CHANG, Y. C., 2010. Multi-variable fuzzy forecasting based on fuzzy clustering and fuzzy rule interpolation techniques. *Information Sciences*, pp. 4772-4783.

CHEN, S. M. & CHEN, C. D., 2011. TAIEX Forecasting Based on Fuzzy Time Series and Fuzzy Variation Groups. *IEEE Transactions on Fuzzy Systems* 19, pp. 1-12.

CHEN, S. M. & CHEN, S. W., 2014. Fuzzy Forecasting Based on Two-Factors Second-Order Fuzzy-Trend Logical Relationship Groups and the Probabilities of Trends of Fuzzy Logical Relationships. *IEEE Transactions on Cybernetics*, pp. 1-5.

CHEN, S. M., CHU, H. P. & SHEU, T. W., 2012. TAIEX Forecasting Using Fuzzy Time Series and Automatically Generated Weights of Multiple Factors. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, Volume 42, pp. 1485--1495.

CHEN, S. M. & KAO, P. Y., 2013. TAIEX forecasting based on fuzzy time series, particle swarm optimization techniques and support vector machines. *Information Sciences* 247, p. 62–71.

D'URSO, P., CAPPELLI, C., LALLO, D. & MASSARI, R., 2013. Clustering of financial time series. *Physica A* 392 , pp. 2114-2129.

D'URSO, P. & MAHARAJ, E. A., 2009. Autocorrelation-based fuzzy clustering of time series. *Fuzzy Sets and Systems*, pp. 3565-3589.

DEVIJVER, P. A. & KITTLER, J., 1982. *Pattern perception; Statistical methods*. 1^a ed. s.l.:Prentice Hall International.

DEZA, M. & DEZA, E., 2009. *Encyclopedia of distances*. s.l.:Springer-Verlag Berlin Heidelberg.

DINGA, X. & LIA, Y. B. A. M. L. P., 2014. An experimental evaluation of novelty detection methods. *Neurocomputing*, Volume 135, pp. 313-327.

EHLERS, R. S., 2009. *Análise de Séries Temporais*. [Online] Available at: <http://www.icmc.usp.br/~ehlers/stemp/stemp.pdf> [Acesso em 17 09 2013].

EUSFLAT, E. S. f. F. L. a. T., 2015. *International Time Series Forecasting Competition - Computational Intelligence in Forecasting*. [Online] Available at: <http://irafm.osu.cz/cif/main.php?c=Static&page=download> [Acesso em 21 03 2015].

FALOUTSOS, C., RANGANATHAN, M. & MANOLOPOULOS, Y., 1994. Fast Subsequence Matching in Time-Series Databases. *Proceedings of the 1994 ACM SIGMOD international conference on Management of data*, pp. 419-129.

FIGUEIREDO, D. B. & SILVA, J. A., 2009. Desvendando os Mistérios do Coeficiente de Correlação de Pearson. *Revista Política Hoje*, Volume 18, n. 8, pp. 115-146.

GALTON, F., 1894. *Natural Inheritance*. 5 ed. New York: Macmillan and Company.

GÓRECKIA, T. & ŁUCZAK, M., 2015. Multivariate time series classification with parametric derivative dynamic time warping. *Expert Systems with Applications* 42, pp. 2305-2312.

GÓRECKI, T. & ŁUCZAK, M., 2014. Non-isometric transforms in time series classification using DTW. *Knowledge-Based Systems* 61, p. 98–108.

GRANGER, C. W. J. & NEWBOLD, J. P., 1976. Forecasting transformed series. *Journal of the Royal Statistical Society*, Volume 38, pp. 189-203.

GRANGER, C. W. J. & NEWBOLD, P., 1977. *Forecasting Economic Time Series*. Academic Press. ed. s.l.:s.n.

HAMMING, R. W., 1950. Error detecting and error correcting codes. *Bell System Technical Journal*, Volume 29, pp. 147-160.

HAUTAMAKI, V., NYKANEN, P. & FRANTI, P., 2008. Time-series clustering by approximate prototypes. *Proceedings of 19th International Conference on Pattern Recognition*, pp. 1-4.

HELENE, O., 2012. *Método dos mínimos quadrados com formalismo matricial*. 2 ed. São Paulo: Livraria da Física.

HENSMAN, J., RATTRAY, M. & LAWRENCE, N. D., 2015. Fast Nonparametric Clustering of Structured Time-Series. *IEEE Transactions on pattern analysis and machine intelligence*, FEBRUARY, Volume 37, pp. 383-393.

ITAKURA, K., 1975. Minimum prediction residual principle applied to speech recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-23, pp. 67-72.

IZAKIANA, H., PEDRYCZ, W. & JAMAL, I., 2015. Fuzzy clustering of time series data using dynamic time warping distance. *Engineering Applications of Artificial Intelligence* 39, p. 235–244.

JANACEK, G. J., BAGNALL, A. J. & POWELL, M., 2005. A likelihood ratio distance measure for the similarity between the Fourier transform of time series. *Lecture notes in computer science*, Volume 3518, p. 737–743.

JENKINS, G. M. & WATTS, D. G., 1968. *Spectral analysis and its applications*. San Francisco: Holden-Day.

JEONG, Y. S., JEONG, M. K. & OMITAOMU, O., 2011. Weighted dynamic time warping for time series classification. *Pattern Recognition* 44, p. 2231–2240.

JI, M., XIE, F. & PING, Y., 2013. A Dynamic Fuzzy Cluster Algorithm for Time Series. *Abstract and Applied Analysis*, pp. 1-7.

JOSHI, B. P. & KUMAR, S., 2013. A computational method for fuzzy time series forecasting based on difference parameters. *International Journal of Modeling, Simulation, and Scientific Computing*, Volume 4, pp. 1250023-1 – 1250023-12.

KENNEY, J. F. & KEEPING, E. S., 1962. *Linear Regression and Correlation*. 3 ed. NJ: Van Nostrand: Princeton.

KEOGH, E., CHU, S., HART, D. & M., P., 2004. Segmenting Time Series: A Survey and Novel Approach. *Data mining in Time Series Databases*, pp. 1-22.

KEOGH, E., CHU, S., HART, D. & PAZZANI, M., 2004. Segmenting time series: A survey and novel approach. *Data mining in time series databases*, Volume 57, pp. 1-22.

KEOGH, E. J., CHU, S., HART, D. & PAZZANI, M. J., 2001. An online algorithm for segmenting time series. *Proceedings of the 2001 IEEE International Conference on Data Mining*, p. 289–296.

KEOGH, E. J. & KASETTY, S., 2002. On the need for time series data mining benchmarks: a survey and empirical demonstration. *International Conference on Knowledge Discovery and Data Mining*, pp. 23-26.

KEOGH, E.; ZHU, Q.; HU, B.; Y., HAO.; XI, X.; WEI, L. & RATANAMAHATANA, C. A., 2011. *The UCR Time Series Classification/Clustering Homepage*. [Online] Available at: http://www.cs.ucr.edu/~eamonn/time_series_data/ [Acesso em 11 10 2014].

KIM, D. W.; LEE, J.; LIM, H.; SEO, J. & KANG, B. Y., 2014. Efficient dynamic time warping for 3D handwriting recognition using gyroscope equipped smartphones. *Expert Systems with Applications*, Volume 41, p. 5180–5189.

KOOPMANS, L. H., 1974. *The spectral analysis of time series*. New York: Academic Press.

KORN, F., JAGADISH, H. V. & FALOUTSOS, C., 1997. Efficiently supporting ad hoc queries in large datasets of time sequences. *Proceedings of the 1997 ACM SIGMOD international conference on Management of data*, pp. 289-300.

KOROL, T., 2014. A fuzzy logic model for forecasting exchange rates. *Knowledge-Based Systems*, Volume 67, pp. 49-60.

KURBALIJA, W., RADOVANOVIC, M., GELERB, Z. & IVANOVICA, M., 2014. The influence of global constraints on similarity measures for time-series databases. *Knowledge-Based Systems*, Volume 56, pp. 49-67.

LEVENSHSTEIN, V. I., 1966. Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *Soviet Physics Doklady*, Volume 10, p. 707–710.

LIAO, T. W., 2005. Clustering of time series data - a survey. *Pattern Recognition* 38, p. 1857–1874.

LI, C. & CHIANG, T. W., 2013. Complex Neurofuzzy ARIMA Forecasting: A New Approach Using Complex Fuzzy Sets. *IEEE Transactions on Fuzzy Systems* 21, pp. 567-584.

LI, H., GUO, C. & QIU, W., 2011. Similarity measure based on piecewise linear approximation and derivative dynamic time warping for time series mining. *Expert Systems with Applications* 38, p. 14732–14743.

LI, H., X., W., LIANG, Y. & GAO, S., 2014. Dynamic time warping based on cubic spline interpolation for time series data mining. *IEEE International Conference on Data Mining Workshop (ICDMW)*, pp. 19-26.

LIU, H. T., WEI, N. C. & YANG, C. G., 2009. Improved time-variant fuzzy time series forecast. *Fuzzy Optimization and Decision Making*, Volume 68, pp. 45-65.

MACQUEEN, J. B., 1967. Some methods for classification and analysis of multivariate observations. *Symposium on Mathematical Statistics and Probability*, Volume 1, p. 281–297.

MAKRIDAKIS, S. & HIBON, M., 1979. Accuracy of forecasting: an empirical investigation. *The Journal of the Royal Statistical Society*, Volume 142, Part 2, pp. 97-145.

MARIOTE, L. E., 2008. *Mineração de séries temporais de dados de sensores*. s.l.:s.n.

MOLLER-LEVET, C. S., KLAWONN, F., CHO, K. H. & WOLKENHAUER, O., 2003. Fuzzy clustering of short time series and unevenly distributed sampling points. *International Symposium on Intelligent Data Analysis*, pp. 28-30.

MORAES, L. A., FLAUZINO, R. A., ARAÚJO, M. A. & BATISTA, O. E., 2013. A Fuzzy Methodology to Improve Time Series Forecast of Power Demand in Distribution Systems. *2013 IEEE Power and Energy Society General Meeting (PES)*, pp. 1-5.

MORETTIN, P. A., 1979. *Análise harmônica de processos estocásticos*. Rio de Janeiro: IMPA-CNPq.

MORETTIN, P. A. & TOLOI, C. M. C., 1981. *Modelos para previsão de séries temporais*. Rio de Janeiro: IMPA.

MORETTIN, P. A. & TOLOI, C. M. C., 2004. *Análise de Séries Temporais*. São Paulo: Edgard Blücher Ltda.

NEWBOLD, P., 1994. *Statistics for Business & Economics*. New Jersey: Printice Hall.

NIEATTRAKUL, V. & RATANAMAHATANA, C., 2007. On clustering multimedia time series data using K-means and dynamic time warping. *IEEE International Conference on Multimedia and Ubiquitous Engineering*, pp. 733-738.

OLIVEIRA, A. L. I. & MEIRA, S. R. L., 2006. Detecting novelties in time series through neural networks forecasting with robust confidence intervals. *Neurocomputing*, Volume 70, pp. 79-92.

PARK, S., LEE, D. & CHU, W. W., 1999. Fast retrieval of similar subsequence in long sequence databases. *Proceedings of the 3rd IEEE Knowledge and Data Engineering Exchange Workshop*, pp. 60-67.

PEDRYCZ, W. & GOMIDE, F., 1998. *An Introduction to Fuzzy Sets: Analysis and Design*. Cambridge, USA: MIT Press.

PETERNELLI, L. A., 2004. *Regressão Linear e Correlação*. [Online] Available at: <http://www.dpi.ufv.br/~peterneli/inf162.www.16032004/materiais/CAPITULO9.pdf> [Acesso em 24 01 2015].

PIMENTEL, M. A. F., CLIFTON, D. A., CLIFTON, L. & TARASSENKO, L., 2014. A review of novelty detection. *Signal Processing*, Volume 99, pp. 215-249.

PUCCIARELLI, A. J., 2005. *Modelagem de Séries Temporais Discretas Utilizando Modelo Nebuloso Takagi-Sugeno*, Campinas. Tese: s.n.

QIU, W., LIU, X. & LI, H., 2011. A generalized method for forecasting based on fuzzy time series. *Expert Systems with Applications*, Volume 38, p. 10446–10453.

RATANAMAHATANA, C. A.; LIN, J.; GUNOPULOS, D.; KEOGH, E.; VLACHOS, M. & DAS, G., 2010. Mining time series data. *Data mining and knowledge discovery handbook*, pp. 1049-1077.

REDDY, D. & PRASANTA, K. J., 2012. Initialization for K-means Clustering using Voronoi Diagram. *International Conference on Computer, Communication, Control and Information Technology*, pp. 25-26.

RODRIGUES, S. C. A., 2012. *Modelo de Regressão Linear e suas Aplicações*, Covilhã - Portugal: s.n.

SAKOE, H. & CHIBA, S., 1978. Dynamic Programming Algorithm Optimization for Spoken Word Recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Volume 26, p. 43–49..

SANTOS, F. J. J. & CAMARGO, H. A., 2013. Preprocessing in Fuzzy Time Series to Improve the Forecasting Accuracy. *12th International Conference on Machine Learning and Applications (ICMLA)*, Volume 2, pp. 170-173.

SANTOS, F. J. J. & CAMARGO, H. A., 2014a. Forecasting in Fuzzy Time Series by an Extension of Simple Exponential Smoothing. *Proceeding of IBERAMIA, LNAI 8864*, p. 257–268.

SANTOS, F. J. J. & CAMARGO, H. A., 2014b. Séries Temporais Fuzzy: Um Modelo de Previsão Baseado no Intervalo Temporal das Amostras. *Proceedings of III Brazilian Congress on Fuzzy Systems (III CBSF)*, pp. 265-276.

SANTOS, F. J. J. & CAMARGO, H. A., NO PRELO. A hybrid forecast model combining fuzzy time series, linear regression and a new smoothing technique. *International Fuzzy Systems Association 2015 (IFSA 2015)*.

SONG, Q. & CHISSOM, B. S., 1993a. Fuzzy Time Series and its Models. *Fuzzy Sets and Systems*, Volume 54, pp. 269-277.

SONG, Q. & CHISSOM, B. S., 1993b. Forecasting Enrollments With Fuzzy Time Series – Part I. *Fuzzy Sets and Systems*, Volume 54, pp. 1-9.

SONG, Q. & CHISSOM, B. S., 1994. Forecasting Enrollments With Fuzzy Time Series – Part II. *Fuzzy Sets and Systems*, Volume 62, pp. 1-8.

SOUZA, R. C., 1989. *Modelos estruturais para previsão de séries temporais: Abordagens clássica e bayesiana*. s.l.:IMPA.

STEPNICKA, M., CORTEZ, P., DONATE, J. P. & STEPNICKOVÁ, L., 2013. Forecasting seasonal time series with computational intelligence: On recent methods and the potential of their combinations. *Expert Systems with Applications* 40, pp. 1981-1992.

SU, C. J., CHIANG, C. Y. & HUANG, J. Y., 2014. Kinect-enabled home-based rehabilitation system using Dynamic Time Warping and fuzzy logic. *Applied Soft Computing*, Volume 22, p. 652–666.

TANUWIJAYA, K. & CHEN, S. M., 2009a. Taiex Forecasting Based on Fuzzy Time Series and Clustering Techniques. *International Conference on Machine Learning and Cybernetics*, pp. 2982-2986.

TANUWIJAYA, K. & CHEN, S. M., 2009b. A new method to forecast enrollments using fuzzy time series and clustering techniques. *Proc. of the Eighth International Conference on Machine Learning and Cybernetics*, pp. 12-25.

THEODORIDIS, S. & KOUTROUMBAS, K., 2009. *Pattern Recognition*. 4 ed. s.l.:Academic Press - Elsevier.

THINH, V. B. & ANH, D. T., 2012. Time Series Clustering Based on I-k-Means and Multi-resolution PLA Transform. *IEEE Research, Innovation, and Vision for the Future*, pp. 1-4.

TORMENE, P., GIORGINO, T., QUAGLINI, S. & STEFANELLI, M., 2009. Matching incomplete time series with dynamic time warping: an algorithm and an application to post-stroke rehabilitation. *Artificial Intelligence in Medicine*, Volume 45, p. 11–34.

USLU, V. R., BAS, E., YOLCU, U. & EGRIOGLU, E., 2014. A fuzzy time series approach based on weights determined by the number of recurrences of fuzzy relations. *Swarm and Evolutionary Computation* 15, p. 19–26.

WANG, C. R. & LEE, S. J., 2014. Temporal Prediction Using Self-Organizing Multilayer Perceptron. *Proceedings of the 2014 International Conference on Machine Learning and Cybernetics*, pp. 585-591.

XIE, Y., WULAMU, Y., WANG, Y. & LIU, Z., 2014. Implementation of Time Series Data Clustering Based on SVD for Stock Data Analysis on Hadoop Platform. *IEEE 9th Conference on Industrial Electronics and Applications (ICIEA)*, pp. 2007-2010.

XI, X.; KEOGH, E.; SHELTON, C.; WEI, L. & RATANAMAHATANA, C. A., 2006. Fast time series classification using numerosity reduction. *Proceedings of the 23rd international conference on machine learning*, pp. 1033-1040.

YAGER, R. R., 2013. Exponential smoothing with credibility weighted observations. *Information Sciences* 252, pp. 96-105.

YAHOO! FINANCE, 2013. *TWII Historical Prices*. [Online] Available at: <http://finance.yahoo.com/q/hp?s=^TWII+Historical+Prices> [Acesso em 25 03 2015].

YEH, M. Y., DAI, B. R. & CHEN, M. S., 2007. Clustering over multiple evolving streams by events and correlations. *IEEE Transactions on Knowledge and Data Engineering*, Volume 19, p. 1349–1362.

YU, D.; YU, X.; HU, Q.; LIU, J. & WU, A., 2011. Dynamic time warping constraint learning for large margin nearest neighbor classification. *Information Sciences*, Volume 181, p. 2787–2796.

YU, T. H. K. & HUARNG, K. H., 2008. A Bivariate Fuzzy Time Series Model to Forecast the TAIEX. *Expert Systems With Applications*, Volume 34, pp. 2945-2952.

ZADEH, L. A., 1965. Fuzzy sets. *Information and Control*, Volume 8, n. 3, pp. 338-353.

ZHANG, G. P., 2003. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* 50, p. 159–175.

ZHANG, X., J., L., DU, Y. & LV, T., 2011. A novel clustering method on time series data. *Expert Systems with Applications*, Volume 38, p. 11891–11900..

Apêndice A

TERMOS LINGUÍSTICOS UTILIZADOS NA ANÁLISE COMPORTAMENTAL DAS ST ECG200 (KEOGH, et al., 2011)

Termo linguístico	Parâmetros dos Conjuntos Fuzzy			
	a_1	a_2	a_3	a_4
0	-3,61	-2,03	-2,03	-1,78
1	-2,03	-1,78	-1,78	-1,32
2	-1,78	-1,32	-1,32	-0,93
3	-1,32	-0,93	-0,93	-0,65
4	-0,93	-0,65	-0,65	-0,24
5	-0,65	-0,24	-0,24	-0,10
6	-0,24	-0,10	-0,10	-0,01
7	-0,10	-0,01	-0,01	0,04
8	-0,01	0,04	0,04	0,18
9	0,04	0,18	0,18	0,20
10	0,18	0,20	0,20	0,24
11	0,20	0,24	0,24	0,30
12	0,24	0,30	0,30	0,33
13	0,30	0,33	0,33	0,40
14	0,33	0,40	0,40	0,42
15	0,40	0,42	0,42	0,42
16	0,42	0,42	0,42	0,43
17	0,42	0,43	0,43	0,43
18	0,43	0,43	0,43	0,44
19	0,43	0,44	0,44	0,45
20	0,44	0,45	0,45	0,47
21	0,45	0,47	0,47	0,48
22	0,47	0,48	0,48	0,50

CONTINUA...

Termo linguístico	Parâmetros dos Conjuntos Fuzzy			
	a_1	a_2	a_3	a_4
23	0,48	0,50	0,50	0,53
24	0,50	0,53	0,53	0,54
25	0,53	0,54	0,54	0,56
26	0,54	0,56	0,56	0,57
27	0,56	0,57	0,57	0,58
28	0,57	0,58	0,58	0,60
29	0,58	0,60	0,60	0,60
30	0,60	0,60	0,60	0,64
31	0,60	0,64	0,64	0,66
32	0,64	0,66	0,66	0,68
33	0,66	0,68	0,68	0,69
34	0,68	0,69	0,69	0,78
35	0,69	0,78	0,78	0,80
36	0,78	0,80	0,80	0,83
37	0,80	0,83	0,83	0,87
38	0,83	0,87	0,87	0,95
39	0,87	0,95	0,95	0,99
40	0,95	0,99	0,99	1,16
41	0,99	1,16	1,16	1,20
42	1,16	1,20	1,20	1,26
43	1,20	1,26	1,26	1,74
44	1,26	1,74	1,74	1,94
45	1,74	1,94	1,94	2,26
46	1,94	2,26	2,26	5,19