

## PREVISÃO EM SÉRIES TEMPORAIS COMBINANDO MODELOS ESTATÍSTICOS E CONCEITOS FUZZY

## FORECASTING IN TIME SERIES COMBINING STATISTICAL MODELS AND FUZZY CONCEPTS

Data de entrega dos originais à  
redação em: 29/02/2016.  
e recebido para diagramação em:  
30/11/2016.

Dr. Fábio José Justo dos Santos<sup>1</sup>  
Dra. Heloisa de Arruda Camargo<sup>2</sup>  
João Carlos Martins Alves<sup>3</sup>

*Os bons resultados obtidos pelas abordagens fuzzy utilizadas para a previsão de valores em séries temporais (ST) tem contribuído significativamente para o crescimento da área. Embora a área estatística também apresente resultados expressivos em relação a previsão de valores, há uma carência de modelos na literatura que combinem séries temporais fuzzy (STF) com modelos clássicos da área da estatística. Assim, este artigo tem como propósito apresentar três modelos fuzzy de previsão que combinam os conceitos das STF e da área estatística para obter uma melhora na acurácia das previsões. Com o objetivo de identificar os pontos fortes e fracos dos modelos e também para validar as abordagens apresentadas, foram realizados experimentos com quatro ST dos índices TAIEX entre os anos de 2001 e 2004. Os resultados obtidos pelos modelos apresentados foram comparados com outras três abordagens disponíveis na literatura.*

**Palavras-chave:** *Séries Temporais Fuzzy. Regressão Linear. Alisamento Exponencial Simples.*

*The good results obtained by fuzzy approaches applied for forecasting values in time series (TS) has contributed significantly to increase the interest in the area. Although also there are expressive results in the statistical, there is a lack of prediction models that combine fuzzy time series (FTS) with classic models of statistics. Thus, this paper aims to present three fuzzy forecasting models that combine the concepts FTS with statistical models to obtain an improvement in the accuracy of forecasts. In order to identify the strong and weak points of the models and to validate the approaches presented, experiments were performed using four TAIEX index between 2001 and 2004. The results obtained were compared with other three approaches available in the literature.*

**Keywords:** *Fuzzy Time Series. Linear Regression. Simple Exponential Smoothing.*

<sup>1</sup>Instituto Federal de Educação, Ciência e Tecnologia do Estado de São Paulo, Câmpus Araraquara.

<sup>2</sup>Departamento de Computação, Universidade Federal de São Carlos

<sup>3</sup>Aluno de Iniciação Científica do curso Tecnólogo em Análise e Desenvolvimento de Sistemas, IFSP Câmpus Araraquara.

## 1 INTRODUÇÃO

Os modelos de previsão desenvolvidos recentemente têm apresentado importantes resultados em aplicações da vida real. Dentre os modelos de previsão disponíveis na literatura, Séries Temporais Fuzzy (STF) (SONG & CHISSOM, 1993) tem sido amplamente utilizados em situações onde existem poucas amostras disponíveis para o treinamento do modelo, ou mesmo onde os dados são representados por meio de termos linguísticos (TL), vagos ou imprecisos.

Diversas abordagens têm contribuído para a evolução da área e, conseqüentemente, para a análise de problemas nos mais diversos cenários do mundo real. Entretanto, dentre as possíveis aplicações de Séries Temporais (ST), certamente a área financeira é a mais explorada em relação aos experimentos para previsão e, por esta razão, a que apresenta maior diversidade de modelos preditivos (CHEN & KAO, 2013) (YU & HUARNNG, 2008) (WANG & LEE, 2014) e (ZHANG, 2003). Assim, este artigo tem como objetivo apresentar a evolução e as principais características dos modelos de pré-processamento e de previsão desenvolvidos a partir da técnica de pré-processamento introduzida em Santos e Camargo (2013). A Seção 2 apresenta os conceitos de STF e as principais etapas presente no cálculo de um valor previsto. A Seção 3 apresenta os modelos de previsão e na Seção 4 os experimentos e uma análise comparativa dos resultados são apresentados. Por fim, as conclusões e trabalhos futuros são descritos na Seção 5.

## 2 SÉRIES TEMPORAIS FUZZY

Séries Temporais Fuzzy (SONG & CHISSOM, 1993) podem ser definidas como uma seqüência de termos linguísticos coletados em um mesmo intervalo temporal onde cada amostra fuzzy representa um valor numérico observado no instante  $t$ . Na seqüência são apresentadas algumas definições sobre STF.

**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 série temporal fuzzy de  $Y(t)$ .

**Definição 2:** 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 lógica fuzzy de primeira ordem.

**Definição 3:** Se  $F(t)$  é causado por  $F(t - 1), F(t - 2), \dots, F(t - k)$ , então esta relação lógica fuzzy (RLF) é definida como relação de implicação de  $k$  -ésima ordem em uma ST, e deve ser representada por  $F(t - k), \dots, F(t - 2), F(t - 1) \rightarrow F(t)$ .

Os procedimentos base apresentados por Song e Chissom (1993) para gerar uma previsão utilizando STF são: (1) Definir o domínio da variável  $Y$ ; (2) Definir os conjuntos fuzzy dentro do domínio para esta variável; (3) Fuzzificar os dados observados na ST; (4) Construir a base de RLF a partir da ST fuzzificada; (5) Gerar uma previsão fuzzificada; (6) Defuzzificar a previsão calculada anteriormente. A fuzzificação a ser realizada na etapa 3, consiste em converter cada amostra numérica da ST em seu conjunto fuzzy de maior pertinência.

## 3 MODELOS FUZZY DE PREVISÃO

Esta seção tem como objetivo apresentar as características e a evolução dos modelos de previsão fuzzy desenvolvidos a partir do método de pré-processamento publicado em Santos e Camargo (2013). Os modelos descritos são compostos por um método de pré-processamento e pelo modelo de previsão.

A etapa de pré-processamento tem como objetivo melhorar a acurácia das previsões por meio da definição do domínio, da identificação e remoção dos *outliers* e por meio da definição do número de TL com suas respectivas funções de pertinência que melhor represente os dados observados. Os modelos são descritos na sequência.

### 3.1 Modelo de previsão em STF baseado no intervalo temporal das amostras

A abordagem introduzida nesta seção tem como principal característica classificar o conhecimento extraído da ST em diferentes grupos temporais (SANTOS & CAMARGO, 2014a).

Concluído o pré-processamento onde os conjuntos fuzzy são definidos, a fuzzificação deve ser realizada e, na sequência, a base de conhecimento deve ser definida por meio de RLF de segunda ordem no formato  $L_i, L_{i+1} \rightarrow L_{i+2}$ , para  $i = 1, 2, \dots, n - 2$ , onde  $n$  é o número de amostras observadas na STF e  $L_i$  é o termo linguístico que representa o valor observado no instante  $i$ . A base de conhecimento é dividida em três grupos temporais de RLF, sendo que o conhecimento mais recente extraído da ST fuzzificada terá maior influência no cálculo do valor previsto do que o conhecimento mais antigo. O grau de influência, ou o peso  $p_{G_i}$ , do grupo de RLF  $G_i$ , para  $i = 1, 2$  e  $3$  deve ser definido empiricamente respeitando a restrição  $p_{G_1} < p_{G_2} < p_{G_3}$ . A previsão é realizada conforme apresentado na Equação 1.

$$d_{n+1} = \sum_{i=1}^3 V_{G_i} \times p_{G_i}, \quad (1)$$

onde  $d_{n+1}$  é o valor previsto,  $p_{G_i}$  é o grau de influência do grupo de RLF  $G_i$ , para  $i = 1, 2$  e  $3$  e  $V_{G_i}$  é a média dos parâmetros centrais dos conjuntos triangulares fuzzy que representam cada implicação das RLF presentes no grupo  $G_i$  com os TL  $L_i = L_{n-1}$  e  $L_{i+1} = L_n$  em seus antecedentes, sendo  $L_n$  o TL que representa a amostra mais recente da STF.

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

Assim como na abordagem anterior, o modelo de previsão apresentado nesta seção utiliza o método de pré-processamento baseado em Santos e Camargo (2013), realizando apenas um ajuste no modo como o número de conjuntos fuzzy é definido para a representação dos valores numéricos observados, como apresentado em Santos e Camargo (2014b).

A principal contribuição deste modelo de previsão está na combinação do método estatístico Alisamento Exponencial Simples (AES) com os conceitos das STF. Este modelo de previsão também introduz o uso de RLF de segunda ordem com três termos linguísticos na implicação, ou seja,  $L_i, L_{i+1} \rightarrow L_{i+2}, L_{i+3}, L_{i+4}$ , para  $i = 1, 2, 3, \dots, n - 4$ , onde  $n$  é o número de amostras observadas (SANTOS & CAMARGO, 2014b). O valor previsto é calculado a partir da Equação 2.

$$d_{n+1} = \alpha_1 \sum_{i=0}^{q-1} (1 - \alpha_1)^i S_{q-i} + (1 - \alpha_1)^q S_0 \quad (2)$$

onde  $d_{n+1}$  é o valor previsto,  $q$  é o número de RLF na base de conhecimento com os antecedentes  $L_i = L_{n-1}$  e  $L_{i+1} = L_n$ , e  $0 \leq \alpha_1 \leq 1$  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,  $\alpha_1$  deve ser definido próximo de 1. Caso deseje-se atribuir maior influência no cálculo do valor previsto ao comportamento mais antigo da ST  $\alpha_1$  deve ser definido próximo a 0. Os valores  $S$  são calculados a partir da Equação 3, sendo  $S_0$  o valor calculado a partir da RLF mais antiga da base de conhecimento com os antecedentes  $L_i = L_{n-1}$  e  $L_{i+1} = L_n$ .

$$S = (1 - \alpha_2)^2 l_{i+2} + \alpha_2(1 - \alpha_1) l_{i+3} + \alpha_2 l_{i+4} \quad (3)$$

onde  $\alpha_2$  indica o fator de suavização e deve respeitar a mesma restrição da Equação 2 e  $l_{i+2}$ ,  $l_{i+3}$ ,  $l_{i+4}$  representam os valores numéricos dos TL das implicações, ou seja, o parâmetro central dos respectivos conjuntos fuzzy triangulares na implicação das RLF com os termos  $L_i = L_{n-1}$  e  $L_{i+1} = L_n$ . Caso não haja ao menos uma RLF na base de conhecimento com os TL  $L_i = L_{n-1}$  e  $L_{i+1} = L_n$ , o valor fornecido como previsão é o parâmetro central do conjunto fuzzy triangular do TL  $L_n$ .

### 3.3 Modelo híbrido de previsão combinando STF, regressão linear e uma nova técnica de suavização

O modelo de previsão apresentado nesta seção bem como sua técnica de pré-processamento foi introduzido em Santos e Camargo (2015). Um dos ajustes que permitiram uma melhor acurácia deste modelo é em relação à técnica de suavização utilizada nas Equações 2 e 3. A nova técnica de suavização é apresentada na Equação 4.

$$w_i = \frac{\theta^i}{\sum_{j=0}^{n-1} \theta^j} \quad (4)$$

onde  $w_i$  é o grau de suavização utilizado para ponderar os  $S_q$  valores na Equação 2 e os valores  $l_i$  na Equação 3,  $n$  é o número de TL a serem suavizados e  $\theta$  é o fator de suavização. Caso  $0 < \theta < 1$  o comportamento mais recente da ST terá maior grau de influência no cálculo do valor previsto e, se  $\theta > 1$ , os primeiros termos disponíveis na ST terão maior influência no cálculo do valor previsto. Quando  $\theta = 1$ , todos os termos terão o mesmo grau de influência.

Outra importante contribuição desta abordagem, é a combinação da 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 média em  $Y_t$  para cada novo dado observado. Assim, quando não há ao menos RLF disponíveis para o cálculo da previsão, o valor previsto é definido por meio do parâmetro  $\hat{\beta}_1$  somado ao parâmetro central do termo linguístico  $L_n$ .

## 4 EXPERIMENTOS E ANÁLISE DOS RESULTADOS

Os experimentos foram realizados utilizando as ST com os índices TAIEX (*Taiwan Stock Exchange Weighted Index*) entre os anos de 2001 e 2004, o que permitiu uma análise comparativa entre diversas abordagens disponíveis na literatura e também conclusões significativas a respeito dos modelos desenvolvidos. O treinamento dos modelos foi realizado utilizando os dados entre janeiro e outubro de cada ano e os testes foram realizados com os índices de novembro e dezembro dos respectivos anos. A Tabela 1 apresenta a acurácia dos experimentos para cada abordagem.

**Tabela 1 – Raiz do erro médio quadrático das previsões para os índices TAIEX com diferentes abordagens**

<b>Abordagem</b>	<b>TAIEX 2001</b>	<b>TAIEX 2002</b>	<b>TAIEX 2003</b>	<b>TAIEX 2004</b>	<b>Média</b>
Yu e Huarng (2008)	120,0	69,0	52,0	60,0	75,2
Chen e Kao (2013)	114,5	76,8	54,2	58,1	75,9
Wang e Lee (2014)	120,2	69,7	56,3	54,2	75,1
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 (2015)	<b>97,2</b>	<b>63,2</b>	<b>48,9</b>	<b>52,2</b>	<b>65,3</b>

Apesar de apresentar uma acurácia satisfatória Santos e Camargo (2014a) apresenta uma acurácia significativamente inferior a das duas outras abordagens apresentadas nas Seções 3.2 e 3.3. A melhora na acurácia obtida em Santos e Camargo (2014b) pode ser atribuída ao uso do AES como técnica de suavização, ao invés da classificação das RLF em grupos e ao uso de RLF de segunda ordem com 3 consequentes. Já a melhora obtida na abordagem Santos e Camargo (2015) é resultado da combinação da regressão linear simples no auxílio da previsão em situações nas quais a base de conhecimento não possui RLF para o cálculo da previsão e também ao uso da nova técnica de suavização apresentada na Equação 4.

## 5 CONCLUSÕES

A combinação dos sistemas fuzzy com os conceitos básicos de ST tem demonstrado por meio de resultados ser uma área de pesquisa bastante promissora. Os modelos de previsão apresentados neste trabalho obtiveram uma acurácia satisfatória nos experimentos realizados, utilizando como princípio a atribuição de diferentes graus de influência aos dados observados ao longo da série.

Como trabalhos futuros está sendo desenvolvida uma nova técnica de suavização que visa identificar automaticamente os valores de  $\theta$  na Equação 4 que geram a melhor acurácia das previsões. Para isso o modelo utilizará o histórico das previsões já realizadas para identificar os valores de  $\theta$  que geraram o menor erro médio quadrático.

## REFERÊNCIAS

- CHEN, S. M.; KAO, P. Y. **TAIEX forecasting based on fuzzy time series, particle swarm optimization techniques and support vector machines**. Information Sciences 247, p. 62–71, 2013.
- SANTOS, F. J. J.; CAMARGO, H. A. **Preprocessing in Fuzzy Time Series to Improve the Forecasting Accuracy**. 12th International Conference on Machine Learning and Applications (ICMLA), Volume 2, p. 170-173, 2013.
- SANTOS, F. J. J.; CAMARGO, H. A. **Séries Temporais Fuzzy: Um Modelo de Previsão Baseado no Intervalo Temporal das Amostras**. Proceedings of III Brazilian Congress on Fuzzy Systems (III CBSF), p. 265-276, 2014a.
- SANTOS, F. J. J.; CAMARGO, H. A. **Forecasting in Fuzzy Time Series by an Extension of Simple Exponential Smoothing**. Proceeding of IBERAMIA, LNAI 8864, p. 257–268, 2014b.

SANTOS, F. J. J.; CAMARGO, H. A. **A hybrid forecast model combining fuzzy time series, linear regression and a new smoothing technique.** 16th World Congress of the International Fuzzy Systems Association (IFSA), p. 1362–1368, 2015.

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

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

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

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