



III Encontro de Iniciação Científica e Tecnológica
III EnICT
ISSN: 2526-6772
IFSP – Câmpus Araraquara
19 e 20 de Setembro de 2018



Análise Comportamental de Séries Temporais Fuzzy.

Lucca Guimarães Scarpa¹, Fábio José Justo dos Santos²

¹ Graduando em Análise e Desenvolvimento de Sistemas, Bolsista PIBIFSP, IFSP Câmpus Araraquara, lucascarpa@ifsp.edu.br

² Docente, IFSP Câmpus Araraquara, fabiojjs@ifsp.edu.br

Área de conhecimento: Lógicas e Semânticas de Programas

RESUMO: Este trabalho tem como propósito realizar a implementação de um modelo de análise comportamental de séries temporais (ST) com o objetivo de identificar STs com comportamentos semelhantes, assim como expor os resultados obtidos com a metodologia utilizada. Para isso, o método apresentado faz uso da métrica *Dynamic Time Warping* (DTW) em conjunto com lógica *Fuzzy*. Com a combinação das duas técnicas espera-se obter um modelo de análise comportamental similar a DTW, mas que utilize técnicas de lógica *Fuzzy* em sua metodologia de análise comportamental. Este modelo de análise é constituído por duas etapas: (1) pré-processamento; e (2) análise. O pré-processamento é constituído pela normalização dos dados, remoção de valores inconsistentes da ST (*outliers*), definição do domínio e do número de conjuntos fuzzy, bem como os parâmetros de representação destes conjuntos. Na etapa de análise, cada valor de uma ST é matematicamente comparado com todos os outros valores das outras ST por meio da Fuzzy DTW, com o objetivo de identificar o grau de dissimilaridade entre elas. O modelo foi aplicado na análise de ST com a cotação do Euro frente a outras vinte e oito moedas diferentes.

PALAVRAS-CHAVE: dynamic time warping; série temporal; análise comportamental;

INTRODUÇÃO

Análise de dados é uma importante área da ciência da computação. Tal área tem o poder de revelar informações ocultas em meio a grande quantidade de dados que são produzidos por diferentes fontes. Uma das necessidades na análise de dados é a comparação entre séries temporais. Séries temporais (ST) são constituídas por uma sequência de dados que tem uma progressão temporal e podem apresentar padrões de crescimento, diminuição, irregularidade ou ainda, sazonalidade.

Um dos desafios na análise do comportamento de uma ST é realizar sua classificação, ou seja, identificar o quão similar uma ST é de outra. Os experimentos deste trabalho foram realizados com dados de cotação do euro frente a moedas de vinte e oito países no ano de 2010, visando identificar quais moedas possuem comportamento semelhante ao ser considerada a cotação do euro.

Para isso foi utilizada uma combinação dos conceitos de conjuntos fuzzy (ZADEH, 1965) com a métrica *Dynamic Time Warping* (BELLMAN, 1957). Entretanto, antes de realizar o cálculo de dissimilaridade por meio da Fuzzy Dynamic Time Warping (FDTW) (SANTOS, 2015) no contexto da cotação do euro, foi preciso realizar um pré-processamento nos dados. Inicialmente os dados das STs devem ser normalizados, pois busca-se analisar a semelhança entre seus comportamentos e não entre seus valores absolutos. Na sequência é realizado o pré-processamento conforme apresentado em Santos e Camargo (2013). Esta etapa do pré-processamento realiza, inicialmente, a remoção de *outliers*, que são valores que representam ruídos na série e que comprometem a precisão da análise. Em seguida, é definido o número e os parâmetros dos termos linguísticos que representarão os valores observados na série. Este processo é realizado a partir da

clusterização dos valores da série temporal, e os resultados obtidos nesta fase serão utilizados mais tarde pelo processo de análise.

Concluído o pré-processamento deve ser realizada a fuzzificação dos valores observados na ST. A fuzzificação consiste no cálculo da pertinência de um valor numérico da ST em todos os termos linguísticos definidos anteriormente. O termo linguístico de maior pertinência representará o valor numérico observado. Após a fuzzificação das STs, a dissimilaridade entre as séries é definida pela métrica Fuzzy DTW (SANTOS, 2015). Ao final, os menores resultados na análise comparativa indicam maior similaridade entre duas STs.

Desta forma, este trabalho tem como objetivo implementar uma métrica de análise de séries temporais que combina *Dynamic Time Warping* (Bellman, 1957) com conjuntos Fuzzy (ZADEH 1965) visando a obtenção de resultados que indiquem a semelhança entre duas ou mais ST. A motivação para o desenvolvimento da pesquisa é a carência de trabalhos semelhantes que implementam tais metodologias de análise de STs e a hipótese da obtenção de bons resultados.

Este artigo está organizado como segue. Na Fundamentação Teórica são apresentados os conceitos dos métodos nos quais o desenvolvimento deste artigo se baseia. Em Metodologia é descrito como ocorreram as etapas do desenvolvimento. Por fim, em Resultados e Discussão são expostos os resultados que foram obtidos por meio da análise de séries temporais dos dados financeiros.

FUNDAMENTAÇÃO TEÓRICA

Este trabalho baseia-se em diversos métodos de análise de dados para que uma técnica de análise comportamental de séries temporais seja implementada com o objetivo de identificar STs semelhantes e obter resultados com uma acurácia satisfatória na análise.

Uma das técnicas é a lógica fuzzy (ZADEH 1965). Neste trabalho sua contribuição consiste em calcular a pertinência de um valor observado na ST nos diversos conjuntos previamente estabelecidos, denominados conjuntos fuzzy ou termos linguísticos. Uma representação possível de conjunto fuzzy é apresentado na Figura 1.

$$A(x) = \begin{cases} 0, & \text{se } x \leq a \\ \frac{x - a}{m - a}, & \text{se } x \in (a, m] \\ \frac{b - x}{b - m}, & \text{se } x \in (m, b) \\ 0, & \text{se } x \geq b \end{cases}$$

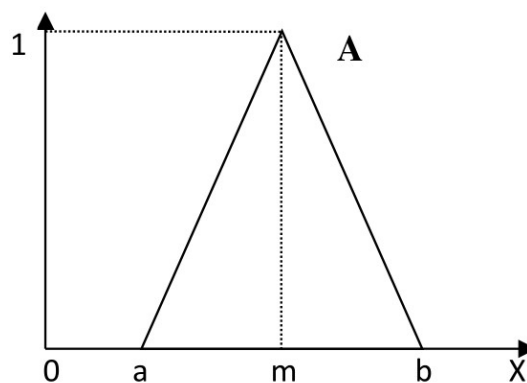


Figura 1: Representação de um conjunto fuzzy triangular
Fonte: SANTOS, 2013

Como apresentado na Figura 1, um conjunto fuzzy triangular é composto por três valores que denotam um intervalo, representados na figura por a , m , b onde a e b são as extremidades e m é o centro. Caso o valor a ser fuzzyficado esteja dentro das extremidades é calculado o ponto de intersecção com a linha, e sua altura representa o grau de pertinência indicado no eixo y , onde 0 representa nenhuma pertinência, ou seja, o valor está fora dos limites de a e b , e 1 indica pertinência completa, que significa que o valor fuzzificado é igual a m .

O pré-processamento de séries temporais é uma etapa crucial para a análise comportamental por meio da Fuzzy DTW, pois é nela que são definidos os termos linguísticos que representarão os valores observados na série. De acordo com Santos e Camargo (2013), o pré-processamento tem início na remoção de *outliers*, ou

seja, ruídos que tem o potencial de reduzir a acurácia dos resultados. Para identificar e remover *outliers* é realizado em cada um dos valores um cálculo para identificar quais valores excedem a tolerância de desvio previamente estabelecida.

Após a remoção de *outliers* é realizada a definição do domínio e dos termos linguísticos que representarão os dados da série. Este processo é realizado a partir do agrupamento de valores semelhantes em *clusters* que são iterados múltiplas vezes até que o centro de todos os clusters esteja nivelado a um *threshold* definido previamente, resultando em grupos cujas extremidades e centro representam os valores de um conjunto fuzzy, conforme apresentado na Figura 1.

Uma vez terminado o pré-processamento de ambas as séries temporais escolhidas tem-se início o processo de análise comportamental. Cada valor de uma ST A é comparado com todos os valores de uma ST B. Essa comparação é realizada por meio da Fuzzy DTW. Para isso, deve ser definida uma matriz $n \times m$, onde n é o número de dados observados na ST A e m o número de dados da ST B. Para cada célula ij da matriz é calculada a distância euclidiana entre os conjuntos fuzzy que representam os elementos i da ST A e j da ST B.

Por fim, a matriz produzida é utilizada para se obter o caminho de menor custo como na métrica DTW padrão, ou seja, a partir da célula $[0,0]$ deve-se percorrer o caminho até $[n, m]$ tomando a direção em que houver o menor valor e sempre carregando consigo o total acumulado pelo caminho. Este valor representa a dissimilaridade entre duas séries temporais. Quanto maior o valor, maior a dissimilaridade.

METODOLOGIA

Para o desenvolvimento deste trabalho foram realizadas pesquisas bibliográficas sobre os conceitos básicos de lógica fuzzy, com foco especial nas definições de conjuntos fuzzy que são fundamentais para o desenvolvimento do projeto.

Alguns métodos de inferência e de desfuzzyficação foram estudados com o objetivo de complementar os estudos sobre lógica fuzzy. Foram desenvolvidos protótipos de teste para que mais a frente, parte do código já implementado possa ser utilizado para em uma análise com Fuzzy DTW.

Após os estudos de lógica fuzzy iniciaram-se os estudos de pré-processamento de séries temporais, onde o objetivo foi manipular os dados reais observados a fim de prepará-los para a etapa de análise comportamental.

Assim como descrito na fundamentação teórica, o sistema foi desenvolvido em etapas. Nos estudos de pré-processamento, para melhor controle dos resultados, todos os testes foram com dados já conhecidos.

Como deseja-se analisar o comportamento da ST e não seus valores absolutos as séries precisam ser normalizadas. Para isso, é realizada a divisão de cada valor da série pelo seu maior valor, resultando em uma ST que possui valores entre 0 e 1. Caso esta normalização não seja realizada, a análise ocorrerá baseada em valores absolutos, ou seja, mesmo que duas séries tenham o mesmo comportamento, mas com escalas diferentes, elas serão classificadas com um alto índice de dissimilaridade apenas por estarem distantes no plano. Este processo iguala as escalas de todas as séries comparadas sem alterar seus comportamentos.

Para a implementação da análise Fuzzy DTW foi necessário primeiro realizar a implementação de um sistema de análise DTW simples. A principal diferença entre DTW e Fuzzy DTW é o fato de que, ao calcular a dissimilaridade entre dois pontos, o DTW padrão utiliza a métrica Distância Euclidiana, enquanto Fuzzy DTW utiliza um cálculo que envolve os termos linguísticos em que os pontos analisados têm maior relevância. O sistema produzido gerou resultados assim como esperado.

A etapa seguinte a ser estudada é a definição de um método para segmentar séries temporais pois ao se analisar uma ST como um todo, certos padrões de comportamento de um segmento podem acabar sendo ocultados durante a classificação, devido ao fato que apenas são comparadas STs completas quando na realidade, fragmentos da ST podem ter comportamentos diferentes. Em paralelo a isso serão realizados estudos do algoritmo de agrupamento Fuzzy C-Means integrado a métrica Fuzzy DTW.

RESULTADOS E DISCUSSÃO

Foram utilizados nos experimentos os valores diários de diversas moedas em comparação ao Euro. Os dados foram coletados no site investing.com (Investing, 2018), considerando a cotação de diferentes moedas em relação ao Euro de 2 de janeiro à 31 de dezembro de 2010.

Após a normalização e o pré-processamento de todas as STs, a análise por meio da Fuzzy DTW foi realizada conforme descrito anteriormente. Os resultados obtidos são apresentados na Tabela 1. As células destacadas representam os maiores e menores valores, ou seja, as STs que possuem maior e menor dissimilaridade.

Como é possível observar na Figura 2, as séries temporais dos valores normalizados das moedas Dirrã dos Emirados Árabes Unidos (AED) e Dólares Americanos (USD) possuem comportamentos visivelmente semelhantes, apresentando o menor valor de dissimilaridade onde seus pontos na Figura 1 estão quase completamente sobrepostos assim como apontado pelos resultados obtidos pela metodologia de análise que estão destacados na Tabela 1.

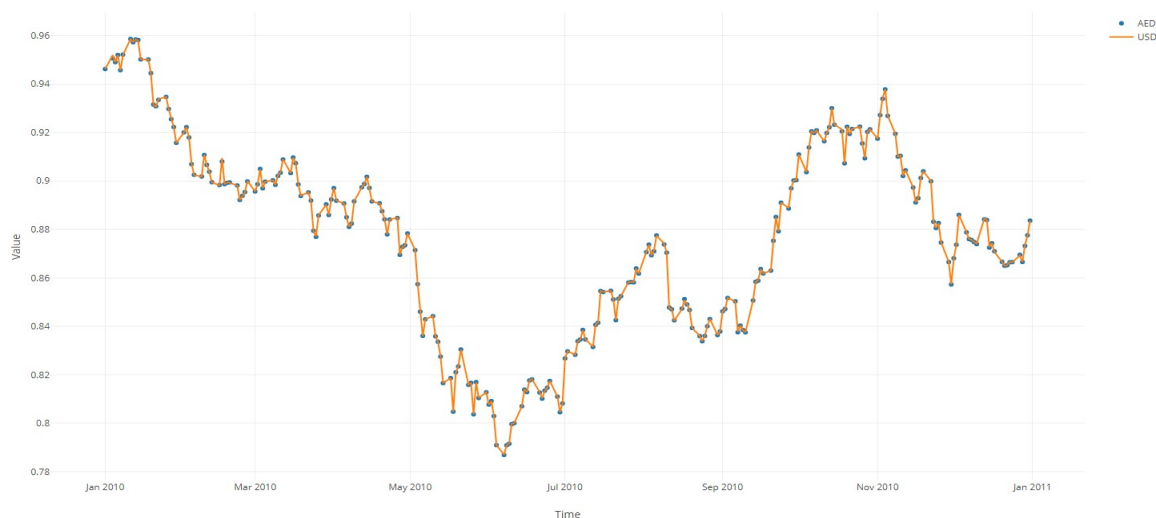


Figura 2: Gráfico de comparação de séries temporais normalizadas mais semelhantes.

EUR /	AED	ARS	AUD	BRL	CAD	CHF	CLP	CNY	CZK	GBP	HKD	HUF	ILS	INR	JPY	KRW	MXN	NOK	PLN	RSD	RUB	SAR	SEK	SGD	THB	TRY	USD	UYU	ZAR	
ZAR	5,8140	5,2764	13,8221	9,5186	7,4373	9,9420	11,0343	7,2160	5,6940	8,6887	5,9281	3,5621	8,1429	8,1734	18,0048	6,5619	8,3933	3,1607	5,9020	1,1362	4,8712	5,8085	11,2649	10,8458	14,6457	6,8578	5,8060	7,3168	-	
UYU	5,7031	4,2128	13,0370	9,1266	6,7275	8,6610	10,2743	6,8131	4,0676	6,0697	5,5238	3,5885	6,9676	7,0894	15,3769	5,4121	7,8762	3,8307	4,5435	1,3235	4,6286	5,6909	8,7621	10,0324	13,7071	4,5898	5,6972	-	-	
USD	6,0858	4,9678	12,5355	9,0423	6,7453	9,0421	9,5512	6,8159	6,5148	7,0430	5,9724	3,8418	6,9297	7,0756	16,1500	6,1297	7,5268	4,4402	5,2757	1,5450	4,8567	6,0688	9,4304	10,6277	14,1006	5,5908	-	-	-	
TRY	6,2599	4,0149	14,1548	11,8304	8,4846	9,2953	12,3774	8,2948	5,0107	3,9015	6,0948	3,0113	7,7420	8,6294	25,4055	6,1199	9,9298	2,7385	4,9562	1,2824	4,4581	6,2608	8,5335	13,6784	22,0310	-	-	-	-	
THB	6,2201	5,4431	14,0893	8,4640	7,6425	11,3687	10,8179	7,4079	6,5041	8,8187	6,2012	4,3407	9,2887	7,4542	14,3896	7,3525	8,2382	4,3627	5,5696	1,4510	7,0218	6,2226	10,3100	9,4526	-	-	-	-	-	
SGD	6,3063	5,6178	13,7771	9,1459	7,6719	9,5442	10,3027	6,8822	6,7259	8,7954	6,4944	4,6095	7,7133	8,5532	17,3679	7,2293	8,8569	4,3241	5,5470	1,3841	4,7394	6,3366	11,0930	-	-	-	-	-	-	
SEK	5,3857	4,1252	15,7880	11,3452	6,5258	8,9702	11,0650	6,4554	4,5499	3,9824	5,2188	4,0944	6,8192	7,3743	28,1168	5,8178	8,5290	4,2557	5,0391	1,9757	3,8667	5,3030	-	-	-	-	-	-	-	
SAR	6,0496	5,0016	12,5279	9,0601	6,7490	9,0337	9,5446	6,7740	6,5295	7,1703	5,9471	3,8380	6,9198	7,0828	16,1542	6,1338	7,5434	4,4627	5,2749	1,5404	4,8880	-	-	-	-	-	-	-	-	-
RUB	5,0049	3,8302	10,5116	8,2292	6,2835	6,8243	8,9877	6,6813	5,5585	4,8569	4,8956	3,6538	5,8493	6,5956	17,1152	5,1847	7,6281	3,2454	5,4764	1,1485	-	-	-	-	-	-	-	-	-	-
RSD	2,1471	2,2279	6,3490	8,2281	7,2708	3,6155	7,2793	6,3286	3,8186	3,0333	2,2224	4,4866	4,9523	7,1426	17,1284	4,8837	8,2716	3,8979	3,5574	-	-	-	-	-	-	-	-	-	-	-
PLN	4,6173	2,8687	12,9257	15,8202	7,3357	8,3997	13,0060	6,8149	2,0517	2,9980	4,4208	2,2894	8,3978	9,7079	24,7714	6,8732	11,1191	2,5274	-	-	-	-	-	-	-	-	-	-	-	-
NOK	5,8427	4,1448	19,8946	23,0117	10,4661	14,3638	20,4231	9,3312	3,2023	5,1838	5,7407	2,7422	13,1611	15,4950	32,0479	11,4634	17,1415	-	-	-	-	-	-	-	-	-	-	-	-	-
MXN	6,1591	5,4309	10,6701	8,5537	6,9289	8,1147	8,9170	7,4598	5,1613	8,9513	6,2462	3,3033	8,4645	7,1282	14,1388	6,7901	-	-	-	-	-	-	-	-	-	-	-	-	-	-
KRW	5,7782	4,4078	10,5686	8,7656	7,3725	7,3414	9,1912	6,8176	5,3763	4,4859	5,6911	3,5284	6,3286	6,8231	15,4397	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
JPY	5,4951	6,6350	11,2912	9,4163	7,0231	9,2256	10,4873	6,4907	5,7736	4,9307	5,7673	4,1041	7,2509	7,5867	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
INR	6,2494	5,2726	11,3247	9,0777	6,8283	8,1388	8,8832	7,4196	6,3569	7,7403	6,1375	3,9161	8,0235	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ILS	5,3718	4,7928	11,5367	10,3673	6,9541	7,6610	9,0273	6,9060	6,1342	4,8984	5,3576	3,9811	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HUF	2,7093	2,2862	9,7278	11,9961	6,4398	5,2262	9,6855	5,9967	3,1315	2,1036	2,5840	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HKD	6,0409	5,0601	12,4884	9,1263	6,8114	8,8024	9,6020	6,8714	6,5374	7,0018	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GBP	6,4385	4,1503	16,9163	14,6208	9,4319	10,8646	14,1209	7,9928	3,2156	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CZK	6,1845	4,3357	21,8559	20,7053	10,0744	16,1710	18,5625	8,9840	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CNY	6,4475	5,1223	12,3477	9,5556	7,6024	8,2376	9,7133	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CLP	4,4011	3,9481	14,6363	9,6434	6,2546	9,9331	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CHF	6,1050	4,7508	19,4231	14,6158	9,1668	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CAD	6,8801	4,9362	11,3295	9,5515	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
BRL	5,3110	4,1162	12,4760	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AUD	5,5481	3,9967	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ARS	5,1719	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AED	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Tabela 1. Relações de dissimilaridade entre séries temporais.

CONCLUSÕES

Após realizadas todas as etapas de desenvolvimento foi implementado com sucesso um método de análise de dados por meio da métrica Fuzzy Dynamic Time Warping, que produziu resultados coerentes com o que se era esperado.

O próximo passo a ser estudado é a definição de um método capaz de realizar análises comportamentais em fragmentos de séries temporais, pois ao se analisar uma ST como um todo, certos padrões de comportamento de um segmento podem acabar sendo ocultados durante a classificação devido ao fato de que apenas são comparadas STs completas, quando na realidade, fragmentos da ST podem ter comportamentos diferentes. Além disso, espera-se integrar o algoritmo Fuzzy C-Means com a métrica Fuzzy DTW, a fim de obter clusters de ST semelhantes.

REFERÊNCIAS

BELLMAN, R. E., 1957. Dynamic Programming. Princeton University Press.

Investing.com, 2018. Disponível em: <https://www.investing.com/currencies>. Acesso em 17/05/2018

SANTOS, F. J. J., 2015 Análise de Séries Temporais Fuzzy para Previsão e Identificação de Padrões Comportamentais Dinâmicos. São Carlos, UFSCar, 132 f. Tese Doutorado. p. 47-50.

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, Miami, FL, 2013, pp. 170-173. doi: 10.1109/ICMLA.2013.185

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.