

Séries Temporais: Analisar o Passado, Predizer o Futuro

Pedro Carvalho de Oliveira
 Departamento de Engenharia Informática
 Universidade de Coimbra
 3030-290 Coimbra, Portugal
 pcoliv@student.dei.uc.pt

Resumo

Pequena síntese sobre a Análise de Séries Temporais, com o objectivo de elucidar sobre os principais modelos computacionais utilizados nessa análise. Evidenciando de uma maneira simples e resumida toda a teoria sobre Séries Temporais, dá-se ênfase á utilização de exemplos reais na explicação da teoria.

Palavras-chave

Time Series Analysis, Time Series Forecasting, Temporal Data Mining, Periodic Patterns Mining.

1. Introdução

A Análise de Séries Temporais é uma importante área de investigação em várias áreas do conhecimento. Neste artigo será feita uma pequena introdução á Análise de Séries Temporais, evidenciando de uma maneira simples e resumida toda a teoria sobre as mesmas. Existem poucos livros e artigos que falem de uma maneira resumida sobre esta temática [9], mas mesmo esses perdem-se em pormenores em alguns aspectos, partindo do princípio que o leitor já possui um relativo conhecimento do tema.

Este artigo foi escrito com base na experiência pessoal do autor em projectos de Análise de Séries Temporais no KDMG (Knowledge Discovery and Management Group) do CISUC (Centro de Informática e Sistemas da Universidade de Coimbra). As referências bibliográficas base foram, respectivamente, [9], [12] e [13].

Numa primeira fase será dada uma base teórico-matemática sobre Séries Temporais, demonstrando a informação que é necessário extrair das mesmas para efectuar esta análise. Serão demonstrados alguns dos modelos computacionais mais básicos utilizados, juntamente com aplicações informáticas que apliquem esses modelos.

2. Séries Temporais

“A time series is a sequence of observations which are ordered in time (or space)” [1]. Apesar de simplista, esta definição de Série Temporal transmite o essencial sobre o assunto. O nosso mundo não é estacionário: muitas variáveis alteram os seus valores através do tempo. Essa

variação é transmitida através de uma série temporal.

Estamos familiarizados a ver Séries Temporais numa representação gráfica (*Imagem 1*), onde habitualmente o eixo das abcissas representa o tempo e o das ordenadas a variação do que estamos a representar. Matematicamente, uma Série Temporal X é representada da seguinte forma:

$$X = \{X_{t-1}, X_{t-2}, X_{t-3}, \dots\}$$

Ou seja, a Série Temporal X corresponde ao conjunto das medições em relação ao tempo t .

Existem dois tipos de Séries Temporais: contínuas, onde temos observações em todos os instantes do tempo (Ex: Electrocardiograma, Bolsa de Valores) e discretas, onde existe uma observação em espaços de tempo normalmente regularem (Ex: Logs de Servidor).

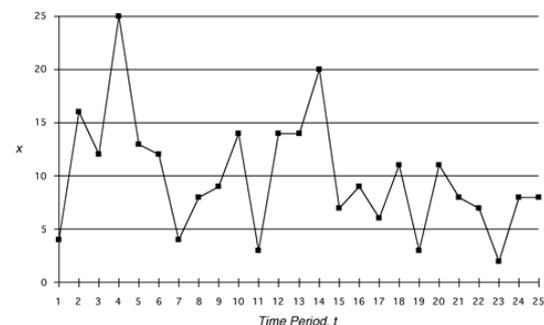


Imagem 1. Série Temporal

3. Análise de Séries Temporais

A Análise de Séries Temporais consiste na aplicação de modelos matemáticos e estatísticos nos dados das mesmas, com o objectivo de quantificar e compreender o fenómeno da

variação temporal. Essa análise é feita com dois objectivos:

- Analisar o passado, tentando retirar conhecimento útil do mesmo
- Predizer o futuro, tentando através da análise dos dados construir um modelo que nos permita facilmente antever a evolução futura da série temporal

3.1 Que informação é importante extrair?

Para efectuar a análise, é necessário identificar e extrair informação relevante da Série Temporal que nos permita atingir os nossos objectivos. Temos de identificar as tendências da Série Temporal, saber quando desceu/cresceu e qual a taxa a que essas modificações ocorreram (Ex: *Imagem 2*). Detectar ciclos que nos permitam identificar padrões temporais (Ex: O número de manchas solares atinge o seu cume a cada 11 anos [2], *Imagem 3*), e descobrir estações, que correspondem a ciclos que ocorrem no espaço de um ano (Ex: O número de vendas de um retalhista aumenta no mês de Dezembro). É também necessário detectar irregularidades e verificar se as mesmas são normais ou foram uma anomalia no registo dos dados (Ex: Um erro no sistema de logging de um servidor). Por ultimo é importante tentar achar relações entre os dados que estamos a analisar e outras variáveis que possam influenciar directamente os mesmos (Ex: A produção de crude no Iraque sofreu 3 grandes baixas nos últimos anos, que correspondem á guerra com o Irão (1981) e com os EUA (1991 e 2003) [3], *Imagem 4*).



Imagem 2. Acções da Sun Microsystems, Inc. em Nasdaq (05/1999 - 01/2001)

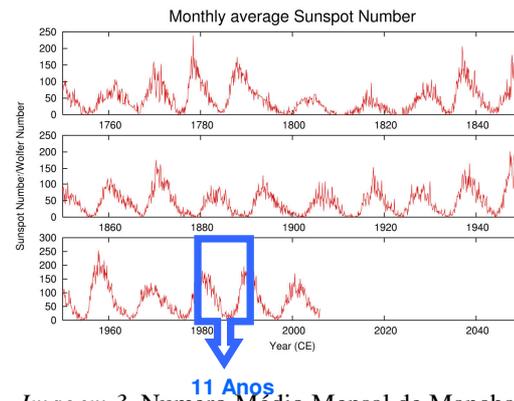


Imagem 3. Numero Médio Mensal de Manchas Solares (1750 – 2005)

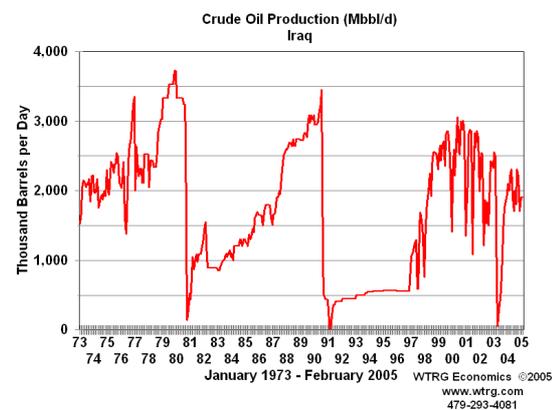


Imagem 4. Produção de Crude no Iraque (1973-2005)

3.2 Modelos utilizados na extracção de informação

Existem 3 modelos matemáticos/estatísticos básicos na extracção de conhecimento de séries temporais.

3.2.1 AR (AutoRegressive) [10]

$$X_t = c + \sum_{i=1}^p \varphi_i X_{t-i} + \varepsilon_t.$$

O valor da Série Temporal X no momento t é baseado numa combinação linear de valores anteriores (backward prediction), futuros (forward prediction) ou ambos (backward-forward prediction). A esse valor normalmente é adicionado um factor de erro ε_t (normalmente uma distribuição normal) e uma constante c inferior a 1.

3.2.2 I (Integrated)

$$(1 - L)^d X_t$$

Normalmente utilizado em conjunto com outros modelos (AR e MA), sendo L o Lag Operator e d o nível de diferenciação.

3.2.3 MA (Moving Average) [11]

Bom a detectar tendências, sendo normalmente utilizado em valores bolsistas e volumes transaccionais. Existem 3 sub modelos distintos:

- **SMA (Simple Moving Average)**

$$SMA = \frac{p_1 + p_2 + \dots + p_n}{n}$$

Corresponde a uma simples média pesada dos valores da Série Temporal. Apesar de simplista, funciona muito bem em dados pouco complexos e periódicos.

- **WMA (Weighted Moving Average)**

$$WMA = \frac{np_1 + (n-1)p_2 + \dots + 2p_{n-1} + p_n}{n + (n-1) + \dots + 2 + 1}$$

Cada valor p da Série Temporal é multiplicado por uma função decrescente, tendo assim cada momento $t-x$ um peso diferente no cálculo do valor da Série no momento t . Desta maneira é dada maior importância aos valores mais próximos no tempo.

- **EMA (Exponential Moving Average)**

$$EMA = \frac{p_1 + (1-\alpha)p_2 + (1-\alpha)^2p_3 + (1-\alpha)^3p_4 + \dots}{1 + (1-\alpha) + (1-\alpha)^2 + (1-\alpha)^3 + \dots}$$

Idêntico ao *WMA*, mas aplicando uma função exponencial decrescente. Desta maneira é dada ainda mais relevância aos valores próximos de t .

Estes modelos normalmente não são aplicados na sua forma base, sendo os modelos mais utilizados usualmente formados por combinações e versões dos mesmos, como são exemplo *ARMA* (AutoRegressive Moving Average) [4], *ARIMA* (AutoRegressive Integrated Moving Average) [5], *ARCH* (Autoregressive Conditional Heteroskedasticity) [6] e *ART* (AutoRegression Trees) [7]. Destes, vamos focar-nos na explicação do último.

3.2.4 ART (AutoRegression Trees) [7]

Este modelo propõe uma generalização do modelo matemático/estatístico *AR* (AutoRegressive), através da aplicação de um algoritmo de Inteligência Artificial: as Árvores de Decisão.

Uma Árvore de Decisão [19] é um modelo preditivo que faz o mapeamento de um item em conclusões sobre o valor provável do mesmo. Esse mapeamento é estruturado em árvore, sendo as folhas os valores prováveis e os restantes nós o esquema de decisão.

Usando uma aproximação *Bayesiana*, uma *ART* consiste numa Árvore de Decisão cujas folhas são constituídas por fórmulas de AutoRegressão (*AR*). Suponhamos que queríamos prever as vendas de um determinado produto, por exemplo

leite, de uma cadeia de supermercados. Aplicando o modelo *ART*, ficaríamos com um modelo idêntico à *Imagem 5*. De realçar a fórmula de AutoRegressão, que demonstra neste caso que a venda de leite é influenciada directamente pela venda de pão.

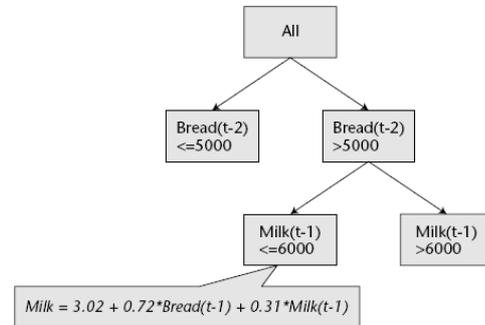


Imagem 5. AutoRegression Tree

3.3 Aplicações Informáticas

A aplicação dos modelos às Séries Temporais, principalmente se o conjunto de dados das mesmas for de grande tamanho, é normalmente efectuada por aplicações informáticas específicas. Algumas das mais usuais são *SAS/ETS* da *SAS* [14], *AutoBox* da Automatic Forecasting Systems [15], *Analysis Services do Microsoft SQL Server 2005* [16] e os open-source *TClass* [17] e *Modified Weka* [18].

4. Conclusões

Numa época em que quase todas as Organizações registam e guardam os seus dados de operação, a análise de Séries Temporais começa a ter um importante papel na definição de metodologias de acção dentro das Organizações, servindo também como instrumento de análise e avaliação de medidas tomadas.

Sendo uma área em grande desenvolvimento, com um crescimento de mercado estimado em 10%-20% ao ano [8], tudo pressagia um futuro promissor para a área.

Referências

- [1] Valerie J.Easton, John H. McColl “Statistics Glossary: Time Series Data”, STEPS, http://www.stats.gla.ac.uk/steps/glossary/time_series.html
- [2] RC Carrington (1863), “Observations of the Spots on the Sun”, Williams and Norgate, London
- [3] Mueller, R. K., Eggert, D. J., Swanson, H. S. (1981) “Petroleum decline analysis using time

- series”, *Energy Economics*, Volume 3, pp. 256-267.
- [4] George Box and F.M. Jenkins (1976) *Time Series Analysis: Forecasting and Control*, Second Edition, Oakland, Holden-Day.
- [5] Mills, Terence C. (1990) *Time Series Techniques for Economists*, Cambridge University Press.
- [6] Robert F. Engle. (2001) "GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics", *Journal of Economic Perspectives* 15(4), pp. 157-168
- [7] C Meek, D Chickering, D Heckerman (2002) “Autoregressive tree models for time-series analysis”, *Proceedings of the Second International SIAM Conference on Data Mining*, pp 229-244.
- [8] METASpectrumSM (2004), *METASpectrum Data Mining Tools Evaluation 2004*, www.metagroup.com.us
- [9] Peter J. Brockwell, Richard A. Davis (1997) *Introduction to Time Series and Forecasting 2nd Edition*, Springer Texts in Statistics
- [10] Wikipedia, “Moving Average (technical analysis)”, http://en.wikipedia.org/wiki/Moving_average
- [11] Wikipedia, “AutoRegressive Moving Average Model”, <http://en.wikipedia.org/wiki/Autoregressive>
- [12] Manuel Filipe Santos, Carla Sousa Azevedo (2005) *Data Mining – Descoberta de Conhecimento em Bases de Dados*, 1ª Edição, FCA
- [13] ZhaoHui Tang, Jamie MacLennan (2005) *Data Mining with SQL Server 2005*, 1st Edition, Wiley Publishing Inc.
- [14] SAS/ETS
www.sas.com/technologies/analytics/forecasting/ets/
- [15] *AutoBox*, www.autobox.com
- [16] *Microsoft SQL Server 2005 – Data Mining*, www.microsoft.com/sql/technologies/dm
- [17] *TClass – A Supervised Learning Algorithm for MultiVariate Time Series*, www.cse.unsw.edu.au/~waleed/tclass
- [18] *Modified Weka*, <http://davis.wpi.edu/~xmdv/weka>
- [19] Wikipedia, “Decision Tree”, http://en.wikipedia.org/wiki/Decision_Tree