

Predição da disponibilidade de recursos computacionais a partir do uso de médias móveis em séries temporais

Prediction of the availability of computational resources from the use of moving averages in time series

RESUMO

Nicolas Charles Oliveira Medeiro
medeiro@alunos.utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Ponta Grossa, Paraná, Brasil

Lourival Aparecido de Góis
gois@utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Ponta Grossa, Paraná, Brasil

Este trabalho tem como objetivo geral a análise de séries temporais geradas através de coletas do percentual de uso de CPU em microcomputadores visando a identificação do comportamento de um equipamento. Para tanto, foi necessário, compreender as características destas séries temporais, e a partir delas, propiciar o estudo e a aplicação de métodos envolvendo médias móveis nos dados coletados a fim de se obter uma assinatura comportamental do equipamento monitorado. Esta assinatura comportamental deverá ao longo de análises sucessivas, ser comparada com o comportamento real do equipamento e com isto, identificar a margem de erro envolvida no processo. A aplicação destes métodos matemáticos possibilitará a modelagem de um sistema de gerenciamento de recursos computacionais (SGR) baseado na previsibilidade de seus índices de ocupação, permitindo e adequando o uso dos mesmos com a maior taxa de sucesso possível.

PALAVRAS-CHAVE: Predição. Série temporal. Recursos computacionais.

ABSTRACT

Recebido: 19 ago. 2020.

Aprovado: 01 out. 2020.

Direito autoral: Este trabalho está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.



This work has as general objective the analysis of time series generated through collections of the percentage of CPU usage in microcomputers aiming at the identification of the equipment behavior. For that, it was necessary to understand the characteristics of these time series, and from them, provide the study and the application of methods involving moving averages in the collected data in order to obtain a behavioral signature of the monitored equipment. This behavioral signature must, throughout successive analyzes, be compared with the actual behavior of the equipment and with this, identify the margin of error involved in the process. The application of these mathematical methods will enable the modeling of a computational resource management system (CRM) based on the predictability of their occupancy rates, allowing and adapting their use with the highest possible success rate.

KEYWORDS: Prediction. Time series. Computational resources.

INTRODUÇÃO

A utilização racional dos recursos computacionais é relevante nos dias atuais, principalmente em um contexto de sistemas distribuídos, como o exemplificado na figura 1, que pode em muitos momentos apresentar equipamentos ociosos, ou seja, estarem realizando somente atividades básicas e com pouca participação nas tarefas do sistema distribuído.

Figura 1: Ilustração de um sistema distribuído



Fonte: *pplware* (2015).

O trabalho de Roesler *et al.* (2018), que em breve síntese consiste no aprimoramento de um sistema MCU (*Multipoint Control Unit*) em nível de *software*, para diminuir o custo financeiro. Na hipótese de haver mais de um MCU disponível no sistema distribuído, é interessante possuir o percentual de uso desse recurso previamente, pois dessa forma é possível evitar que haja sobrecarga em um MCU e deixar outro em estado ocioso.

A ideia de tentar prever acontecimentos futuros também é explorada em outras áreas do conhecimento, como pode ser constatado nos estudos de Silva *et al.* 2008, em que se busca a previsão da temperatura média de Uberlândia – MG, por meio de séries temporais. O artigo de Ehlers (2009) explica alguns modelos preditivos, como por exemplo, o *Holt-Winters*, um modelo de suavização exponencial dos dados.

Se estes estudos estiverem voltados para recursos computacionais, é possível que a partir da coleta de dados de utilização da CPU de cada máquina em um sistema distribuído, gerar séries temporais (percentual de uso de CPU x intervalos pré-definidos) e descobrir como os recursos monitorados destes equipamentos se comportam no passado.

Entretanto, para um SGR, o mais importante é prever como estes comportamentos estarão em um futuro próximo. Partindo desta premissa, uma abordagem largamente aplicada em séries temporais é o uso de médias móveis. Segundo Ehlers (2009), um modelo matemático que realiza a predição, ou seja, a projeção do comportamento dos dados analisados de modo que se possa planejar e escalonar onde uma determinada tarefa será executada no futuro, escolhendo com sucesso uma máquina com capacidade de processamento para tal.

Sendo assim, este trabalho teve como objetivo principal viabilizar um modelo de previsão do comportamento de recursos computacionais através de análises qualitativas de séries temporais obtidas de equipamentos monitorados, buscando com isso, a definição de tendências, sazonalidades e anomalias com a aplicação das médias móveis na definição da assinatura comportamental destes equipamentos e, a comparação desta com o comportamento real dos equipamentos.

MATERIAL E MÉTODOS

Para cumprir com o objetivo proposto, ou seja, analisar o comportamento das séries temporais geradas pelas coletas do percentual de uso de CPU, foram seguidos alguns passos, como pode ser visto a seguir:

- Compreender o funcionamento de séries temporais;
- Verificar o comportamento de um equipamento, através da coleta do percentual de uso de CPU;
- Estudar médias móveis e suas variações;
- Aplicar médias móveis nas séries temporais para definição da assinatura comportamental;
- Analisar a assinatura comportamental em relação ao comportamento real.

Para o início do trabalho, foi necessário compreender o funcionamento de uma série temporal, suas definições e propriedades. Também foi necessário o estudo de um método de predição sendo adotado para este projeto, as médias móveis.

Como definido na figura 2, inicialmente procedeu-se com a captura dos índices de ocupação da CPU nos equipamentos. Para isto foi utilizado a ferramenta nativa das plataformas *Windows* denominado *PERFMON*. A partir da série temporal obtida, o uso das médias móveis possibilitou a geração de uma segunda série, suavizada e contendo somente os dados essenciais para a definição do comportamento do equipamento, neste trabalho denominado como assinatura comportamental. Nesta etapa foram verificadas possíveis sazonalidades, ou seja, trechos da série temporal que se repetem, e principalmente a convergência entre os valores previstos e os reais.

Figura 2: Fluxograma com os passos práticos do trabalho



Fonte: Autoria própria (2020).

Todo o projeto foi dividido em quatro grandes etapas, de modo a facilitar o entendimento e a execução do objetivo principal, como será visto a seguir:

- Levantamento bibliográfico:** Compreender séries temporais, predição e análise da assinatura comportamental (EHLERS, 2009) (GÓIS, 2009) (SILVA, 2016);
- Coleta de dados:** Capturar o percentual de uso de CPU em um conjunto de máquinas (ROESLER et al., 2018);
- Geração da série temporal:** Com os dados coletados, gerar a série temporal de cada máquina para estudo de sua assinatura comportamental (EHLERS, 2009) (SILVA, 2016) (SILVA; GUIMARÃES; TAVARES, 2008);
- Análise da assinatura comportamental:** Verificar sazonalidades e convergência (EHLERS, 2009) (STROBEL, 1987).

Existem algumas variações de médias móveis, para este trabalho utilizou-se Médias Móveis Simples (MMS) ou *Simple Moving Average* (SMA) (EHLERS, 2009). Matematicamente, a média móvel simples nada mais é que uma média de um determinado período de dados dispostos em uma série temporal. Por exemplo, a média de preços de um determinado produto a cada dez dias.

Com o incremento das unidades temporais (horas, dias, meses, anos), no exemplo em questão são dias, novos dados vão substituindo os dados de preços mais antigos, formando uma base amostral para elaboração da média que varia ao longo do tempo, neste caso com uma janela móvel (n em Equação 2) de 10 dias como visto na Equação 1. O comportamento destes preços poderá ser verificado analisando-se a série MMS gerada:

$$MMS = \frac{\text{preço}_0 + \text{preço}_1 + \dots + \text{preço}_9}{10} \quad (1)$$

A equação 1 pode ser definida para ser usada em qualquer situação, sendo assim representada:

$$MMS = \frac{\text{valor}_1 + \text{valor}_2 + \dots + \text{valor}_n}{n} \quad (2)$$

A média móvel simples atribui peso igual para cada componente, ou seja, todos os dados possuem o mesmo peso na média. Tal método é eficaz para valores não discrepantes entre si, pois como se trata de média simples, há uma forte tendência de os valores calculados se aproximarem da média dos valores da janela móvel, desta forma não representando com precisão os valores do conjunto com alto desvio padrão. Neste estudo foi considerado que o comportamento dos equipamentos analisados segue uma tendência de utilização e consequentemente, com anomalias facilmente detectadas.

O método utilizado para a verificação da convergência das séries temporais é o teste de *Wald-Wolfowitz* (STROBEL, 1987), que pode ser assim enunciado: Seja $\{X_t, t = 1, \dots, N\}$ uma série temporal com N observações. Onde M sendo a mediana das N observações de X_t . Atribui-se a cada valor de X_t o símbolo "i" se ele for maior ou igual a M , e "j" caso ele for menor que M . Então, têm-se que $N = ("N_i" \text{ pontos "i"} + ("N_j" \text{ pontos "j"}))$. Desta forma, tem-se grupos de observações marcadas com "i" e grupos de observações marcadas com "j" ao longo da série temporal. O número total de grupos será a estatística do teste, isto é: $K = \text{"número de grupos com símbolos iguais"}$.

Considerando as seguintes hipóteses:

$$\begin{cases} H_0: \text{Existe convergência;} \\ H_1: \text{Não existe convergência.} \end{cases}$$

Rejeita-se a hipótese nula H_0 se houver um número pequeno de grupos com símbolos iguais, ou seja, se K for relativamente pequeno. Para valores de N_i e N_j superiores a 30, pode-se utilizar o Teorema Central do Limite e aproximar a distribuição de K por uma normal, isto é, $K \sim N(\mu, \sigma^2)$ (STROBEL, 1987), sendo:

$$\mu = \frac{2N_i N_j}{N} + 1 \quad (3)$$

$$\sigma = \sqrt{\frac{2N_i N_j (2N_i N_j - N)}{N^2 (N-1)}} \quad (4)$$

As equações 3 e 4 são os passos iniciais para utilizar o modelo de distribuição normal de probabilidade, este modelo, segundo Ehlers (2009), é necessário para quando o espaço amostral, neste caso os valores de N_i e N_j , são superiores a 30.

O resultado após a aplicação deste método pode ser assim interpretado via teste de hipótese:

- Se maior ou igual a 50%: Há convergência;
- Se menor que 50%: Não há convergência.

RESULTADOS E DISCUSSÃO

Para poder demonstrar o resultado, utilizou-se a coleta de um servidor *web* da UTFPR Câmpus Ponta Grossa, esses dados foram utilizados com o objetivo de se compreender os passos para a geração da predição, desde a captura de dados até a análise da série final gerada.

O procedimento para coleta e geração da série temporal e predição é o mesmo para qualquer quantidade de máquinas. Para os gráficos que serão vistos a seguir, utilizou-se o pacote Office® da *Microsoft*, especificamente o *Excel*, pois nele é possível a plotagem de gráficos, além de também disponibilizar funções para cálculo da Média Móvel Simples (MMS).

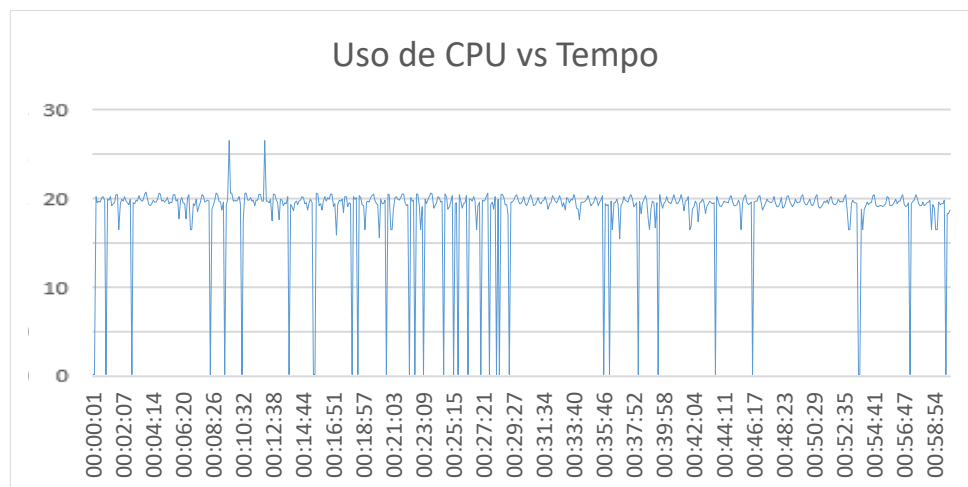
Essa base de dados contém observações do percentual de uso de CPU de um dia todo de uma máquina, desde as 00:00 horas até as 23:59 horas do dia 12/05/2017, com um intervalo de 6 segundos entre cada observação, totalizando 14382 amostras.

Serão apresentados três gráficos:

- a) Coleta;
- b) Predição;
- c) Coleta + predição.

Como há muitos dados na coleta, e para fins de melhor visualização dos dados, aqui serão apresentadas as primeiras 600 amostras, ou seja, uma hora de dados coletados. A janela móvel utilizada na MMS foi de 60 observações, ou seja, 10 minutos. A figura 3 apresenta o gráfico os percentuais observados de coleta de CPU:

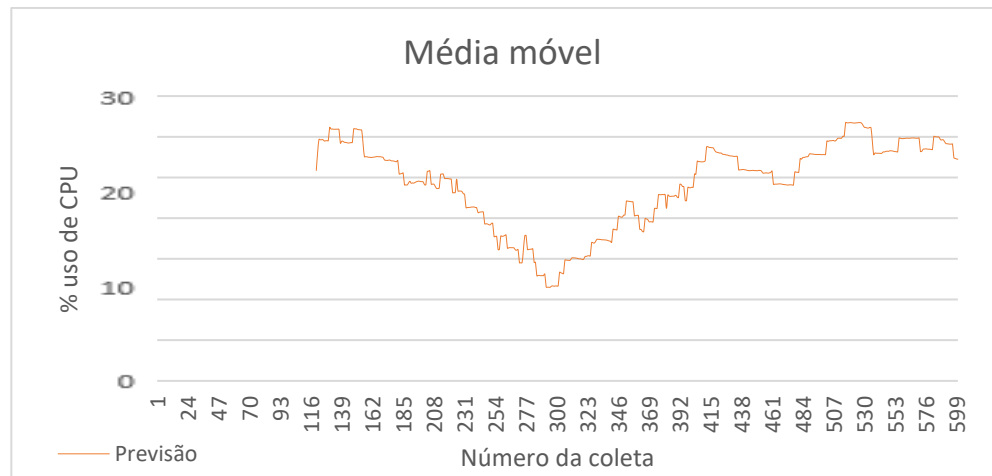
Figura 3: Percentual de uso de CPU em um intervalo de tempo



Fonte: Autoria própria (2020).

Na figura 3 constam somente os dados históricos, ou seja, coletados previamente. De posse disso, foi possível utilizá-los no método preditivo de MMS, através das funções existentes no Excel, conforme descrito na Equação (2) e com os valores obtidos e exibidos na Figura 4, representando o comportamento do equipamento observado:

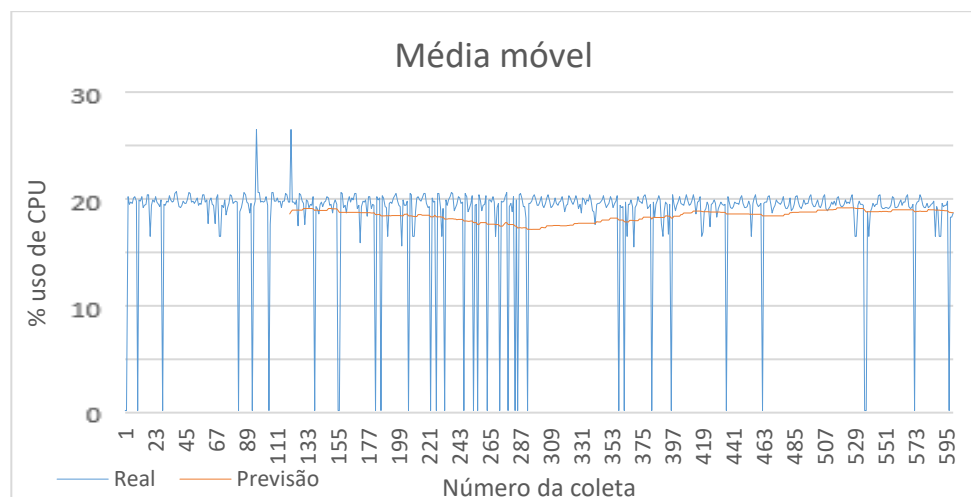
Figura 4: Comportamento a partir da predição do percentual de uso de CPU



Fonte: Autoria própria (2020).

Pode-se observar valores discrepantes no comportamento, isto ocorre devido a série histórica ou temporal possuir alguns valores com alto desvio padrão, e como a MMS é uma média calculada com janelas pré-definidas, neste experimento com 60 observações, ou 10 minutos. A Figura 5 apresenta o gráfico comparativo da série temporal da utilização da CPU e o seu comportamento obtido através da MMS:

Figura 5: Comparativo série histórica vs predição

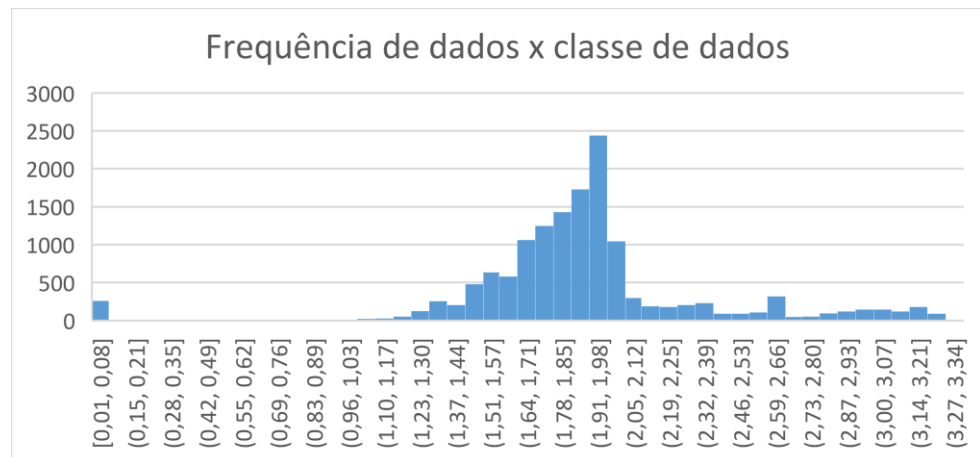


Fonte: Autoria própria (2020).

Com este gráfico comparativo, é possível perceber que a previsão dos dados apresenta uma tendência para onde se concentra a maior parte dos dados, este resultado é esperado, pois está sendo calculada a média dos valores como medida de predição dos dados, como descrito em Ehlers (2009). O teste de convergência descrito por *Wald-Wolfowitz* é realizado no Excel, e como os valores de N_i e N_j são

maiores que 30, deve-se utilizar a distribuição normal para definir a convergência, segundo Ehlers (2009). A Figura 6 ilustra o histograma com as frequências de ocorrências dos percentuais de uso da CPU e com as classes (intervalo) dos dados:

Figura 6: Histograma com as coletas de CPU



Fonte: Autoria própria (2020).

Como a distribuição das coletas da CPU segue o padrão da distribuição normal N , descrita por Ehlers (2009), pode-se aplicar este método para verificação de convergência. Após a aplicação da distribuição normal, no Excel, o valor retornado foi de 0,634831, ou seja, que há aproximadamente 63% de similaridade ou convergência dos dados previstos, ao realizar o teste de hipótese descrito por Strobel (1987) conclui-se que a previsão converge com os dados coletados, portando pode-se concluir que houve êxito ao aplicar médias móveis para prever o uso da CPU no futuro.

CONCLUSÃO

Com a realização deste trabalho, pode-se perceber que foram obtidos com precisão os dados de uso da CPU com a ferramenta escolhida, e após a aplicação de MMS obteve-se a predição, que sua precisão testada e sua acurácia confirmada, demonstrando que os passos anteriores foram concluídos com êxito.

A definição da assinatura comportamental de um equipamento monitorado, permitirá a transferência desta informação para o módulo gerente do Sistema de Gerenciamento de Recursos (SGR), com uma quantidade de dados bem menor do que se este equipamento informasse seus índices de ocupação de CPU a cada 6 segundos, conforme definido neste trabalho. Apesar de não ser objeto deste estudo, fica evidenciado de que a comunicação dentro do SGR irá ocorrer com um overhead bem mais otimizado.

Considerando-se que o SGR terá armazenado em suas bases de dados, somente as assinaturas comportamentais de seus equipamentos gerenciados, e que estas assinaturas representam o conhecimento prévio de como estarão os índices de ocupação dos recursos computacionais em um futuro próximo, a escolha dos mais adequados para a execução de uma tarefa distribuída poderá ser feita com uma taxa de sucesso e com uma confiabilidade aceitável dentro de parâmetros definidos pelo SGR.

Este projeto de iniciação científica apresentou uma abordagem possível e aplicável em um SGR, a qual deverá ser exaustivamente testada e expandida através do prosseguimento das pesquisas, visando a obtenção de resultados mais significativos e comparativos dentre outras abordagens que estão sendo estudadas pelos outros integrantes do Grupo de Pesquisa em Redes e Sistemas Distribuídos.

AGRADECIMENTOS

Agradeço ao meu orientador, Prof. Dr. Lourival Aparecido de Góis pelos apontamentos durante a execução deste trabalho, também agradeço ao Programa Institucional de Voluntariado em Iniciação Científica e Tecnológica (PIVICT) 2019/2020 pela oportunidade de participar de um projeto de iniciação científica.

REFERÊNCIAS

EHLERS, Ricardo Sander. **Análise de séries temporais**. 5. ed. Curitiba: Universidade Federal do Paraná, 2009. 118 p.

GÓIS, Lourival Aparecido de. **Estratégias para comercialização de recursos computacionais em desktop grids**. 2009. 125 f. Tese (Doutorado) - Curso de Engenharia Elétrica e Computação, Departamento de Telemática, Universidade Estadual de Campinas, Campinas, 2009.

ROESLER, Valter; GASPARONI JÚNIOR, Mário; HUSEMANN, Ronaldo; COSTA FILHO, Roberto Irajá T. da; VALLE, Rafael. MCU em software distribuído em nuvem. **Webmedia'2018: Workshops e Pôsteres**, Salvador, p. 127-131, 12 ago. 2020.

SILVA, Cidiney J.. **Predição de Séries Temporais no Contexto de Smart Grids**. 2016. 123 f. Tese (Doutorado) - Curso de Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal de Minas Gerais, Belo Horizonte, 2016.

SILVA, Maria I. S.; GUIMARÃES, Ednaldo C.; TAVARES, Marcelo. Previsão da temperatura média mensal de Uberlândia, MG, com modelos de séries temporais. **Revista Brasileira de Engenharia Agrícola e Ambiental, Campina Grande**, v. 12, n. 5, p. 481-485, 30 jan. 2008.

STROBEL, Orlando Maciel. Homogeneidade de séries temporais. **Revista Tecnologia e Humanismo**, Curitiba, v. 2, n. 3, p. 23-37, set. 1987.