

# GRÁFICOS DE CONTROL DE APRENDIZAGEM DOS PROCESSOS: UMA FLEXIBILIZAÇÃO DE CONTROL ESTADÍSTICO DE PROCESSOS (CEP) COM INTELIGÊNCIA ARTIFICIAL (IA)

Paulo Henrique Ferreira<sup>1</sup>, Laion Lima Boaventura<sup>1</sup>, and Rosemeire  
Leovigildo Fiaccone<sup>1</sup>

<sup>1</sup>Departamento de Estatística, Universidade Federal da Bahia, Salvador, Bahia,  
Brasil, *paulohenri@ufba.br*, *englimaboaventura@gmail.com*, *fiaccone@ufba.br*

## RESUMO

Neste trabalho, técnicas de Inteligência Artificial (IA) e Controle Estatístico de Processos (CEP) são utilizadas para fornecer novas ferramentas para monitoramento de processos. Com base em gráficos de controle mais tradicionais, como propostos por Haworth (1996), bem como apoiados em metodologias mais recentes, como Mostajeran et al. (2016), a Carta de Controle de Classificação (class-chart) e a Carta de Controle de Predição para Respostas Contínuas (pred-chart), aqui propostas, são motivadas principalmente por se apresentarem na literatura como alternativas mais robustas e flexíveis às ferramentas mais tradicionais. Para a iso, além da grande capacidade de reconhecer padrões e diagnosticar problemas, independentemente do cenário de amostragem, essas novas ferramentas são capazes de desempenhar suas funções de monitoramento em larga escala, prevenindo situações de mercado, nas quais os processos ocorrem, cada vez mais . , em grandes volumes de dados. Buscando ilustrar plenamente a flexibilidade dos gráficos propostos, 16 modelos de Aprendizado de Máquina (MA) são aplicados a um grande número de cenários amostrais simulados, variando tamanho amostral, número de amostras, existência ou correlação não linear entre as variáveis preditoras. processo, bem como desequilíbrio e não normalidade da característica de interesse. Além de dois estudos de simulação, a fim de ratificar a aplicabilidade das ferramentas propostas, em relação a outras cartas já consolidadas na literatura, existem 4 aplicações envolvendo bancos de dados reais, cujos

**Palavras-chave:** Classificação, Controle Estatístico de Processos (CEP), Gráficos de Controle, Inteligência Artificial (IA), Predição.

## 1 Introdução

Hoje em dia, a busca contínua das indústrias, dos mais variados setores, por métodos mais precisos de controle da qualidade, está inteiramente ligada ao grande aumento da competitividade entre as empresas. Devido a esse fato, monitorar as variáveis envolvidas no processo de produção, de modo a torná-lo mais eficiente (isto é, reduzir os desperdícios, utilizar sua plena capacidade e, assim, otimizar os resultados), é uma das crescentes preocupações do mercado. De fato, os processos de produção são extremamente importantes para qualquer área de negócio e, portanto, devem ser controlados e analisados, permitindo que as empresas consigam alcançar e monitorar o sucesso almejado. resultados mostram que, tanto uma classe quanto uma carta quântica ou pré-carta, podem ser alternativas interessantes quando se trata de controle de processos.

Pode-se definir um processo como uma atividade que recebe um input (entrada), passa por uma agregação de valor e gera um output (saída). Assim, quando associa-se os processos a modelos de regressão, entende-se essa saída como a variável resposta, ou característica de interesse; e a entrada, como a(s) variável(eis) explicativa(s) utilizada(s) para decifrar o comportamento do valor observado na saída. Entende-se também essa "agregação de valor" como a forma com que as entradas se relacionam entre si para produzir os valores observados na saída, o que no contexto estatístico representa a função (ou modelo) de regressão.

Contudo, modelos de regressão e controle de processos não é uma abordagem nova. Na realidade, desde Mandel [1], propostas vêm sendo elaboradas conectando os conceitos de modelos de regressão e controle de processos, mais precisamente através dos gráficos de controle (ver [2, 3]). Os gráficos (ou cartas) de controle são uma das principais ferramentas do CEP, criados por Walter A. Shewhart, um físico dos laboratórios da Bell Telephone Company, nos anos de 1920. Montgomery [4] aponta que, através do CEP, é possível estabelecer padrões, comparar desempenhos, verificar e estudar desvios, buscar e implementar soluções, analisar novamente o processo após as modificações, buscando a melhor performance de máquinas e/ou pessoas

## 2 Revisão de Literatura

Essa conexão entre modelos de regressão e controle de processos tem acompanhado o avanço tecnológico e muitos trabalhos têm surgido ligando conceitos de IA a controle de processos, a fim de propor novas ferramentas de controle (ver [5–15]; entre outros). Em particular, Sena [15] propõem o uso de gráficos de controle baseados em modelos Generalizados Autorregressivos de Médias Móveis (GARMA), com distribuições de Poisson e Binomial Negativa, para o monitoramento dos casos de malária dos tipos *Plasmodium Vivax*, *Plasmodium Falciparum* e *Plasmodium Mista*, com o intuito de detectar e também prever possíveis surtos ou epidemias em estados da região da Amazônia Legal brasileira. Contudo, a maior parte dos trabalhos relacionados a CEP e IA trata de um dos mais relevantes aspectos do CEP: a análise de padrões não aleatórios. Esses padrões fornecem valiosas informações sobre a melhoria de um processo. O uso de IA sobre tais padrões, visa associar um conjunto específico de causas assinaláveis a um determinado padrão não natural em um gráfico de controle, estreitando, assim, o conjunto de causas que devem ser investigadas em possíveis problemas, otimizando o tempo de diagnóstico e reduzindo substancialmente a probabilidade de uma tomada de decisão equivocada sobre o processo.

## 3 Objetivos e Metodologia

Com base nisso, neste trabalho (originalmente, uma dissertação de mestrado desenvolvida no Programa de Pós-Graduação em Matemática, área de concentração em Estatística, da Universidade Federal da Bahia, pelo aluno Laion Lima Boaventura, sob a supervisão dos professores Rosemeire Leovigildo Fiaccone e Paulo Henrique Ferreira da Silva) são propostas duas novas ferramentas de controle, para situações em que existe o conhecimento prévio acerca dos insumos que fazem parte do processo: o Gráfico de Controle de Classificação (class-chart), isto é, quando a característica de interesse (variável resposta ou de monitoramento) é multicategórica e latente; e o Gráfico de Controle de Predição para Respostas Contínuas (pred-chart), isto é, quando a característica de interesse (variável resposta ou de monitoramento) é mensurável e observável (em geral, contínua). Para a construção dos gráficos propostos, alguns concei-

tos teóricos são necessários, como o método introduzido por Woodall & Montgomery [16] e Vining [3], que se refere à implementação dos gráficos de controle em duas fases; bem como alguns conceitos de Aprendizagem Estatística de Máquina (AEM) e modelos de classificação e regressão. Além disso, ambos os gráficos fazem uso de um parâmetro de tunagem para calibração dos limites, mediante uma política de qualidade adotada (por exemplo, 6 Sigma).

O objetivo é que ambos os gráficos propostos sejam alternativas mais robustas, generalistas e flexíveis do que os gráficos tradicionais (como, por exemplo, o Gráfico de Controle Multinomial, proposto por Amirzadeh [17], para o controle de classificações; e o Gráfico de Controle de Regressão Múltipla, proposto por Vining [3], para o controle de predições de respostas contínuas) utilizados para controle de processos.

## 4 Resultados

Para isso, após apresentar todo o conceito matemático por trás da construção dos gráficos, foi elaborado um minucioso estudo de simulação. Neste, foram desenvolvidos, entre os dois gráficos, 96 diferentes cenários, nos quais variou-se: tamanho amostral, número de variáveis explicativas, comportamento da variável resposta e número de amostras; tudo isso aplicado a 16 diferentes modelos de regressão (Support Vector Regression, Lasso, Ridge, Elastic Net, Regressão Quantílica, Modelos Lineares Generalizados - Distribuições Gaussiana e Gama, SpAM e Regressão Polinomial) e classificação (Support Vector Machine, Árvore de Decisão, K-Vizinhos Mais Próximos, Análise Discriminante Linear e Quadrática, Floresta Aleatória, Rede Neural e Regressão Multinomial). Maiores detalhes sobre essas técnicas podem ser encontrados no livro de Hastie [18]. Os resultados das simulações foram ilustrados sob a medida ARL (do inglês Average Run Length), hipotetizando tanto cenários de conformidade quanto de não conformidade dos processos. Em suma, os resultados obtidos ratificaram os conceitos esperados de vantagens e desvantagens, os quais indicaram como principal vantagem a robustez dos gráficos propostos, independente da característica do modelo de regressão e classificação utilizado, enquanto que, também como esperado, o custo computacional se mostrou um problema, uma vez

que o número de iterações necessárias no processo de construção dos gráficos cresceu conforme o aumento da dimensionalidade amostral. Para que não ficassem lacunas conceituais, tanto na apresentação dos gráficos propostos quanto no desenvolvimento dos estudos de simulação, foi mostrado todo o levantamento bibliográfico referente aos modelos de regressão e classificação, aos principais gráficos de controle já existentes na literatura, bem como aos conceitos de IA, mais precisamente de AM, requeridos na implementação dos gráficos propostos.

Por fim, de modo a ratificar os estudos elaborados, bem como comparar as propostas a outros gráficos já consolidados na literatura, foram apresentados 4 exemplos de possíveis aplicações dos gráficos desenvolvidos. Todos os exemplos/conjuntos de dados utilizados foram reais (três deles em aplicações inéditas, sendo pertencentes a uma grande empresa de varejo da América Latina, e um deles referente a um conjunto de dados de benchmarking da área preditiva, disponibilizado no site Kaggle), o que possibilita o leitor a conectar situações de mercado às ferramentas desenvolvidas. Em suma, os novos gráficos de controle (class-chart e pred-chart) apresentaram limites mais flexíveis do que os gráficos tradicionais (Gráfico de Controle Multinomial e Gráfico de Controle de Regressão Múltipla, respectivamente), o que refletiu em um número menor de alarmes falsos no processo. Além disso, foi possível observar que ambos os gráficos se mostraram eficientes em identificar a mudança de distribuição do processo.

## 5 Conclusões

Os resultados obtidos neste trabalho (exibidos, em detalhes, em [19,20]) apontaram que tanto o class-chart quanto o pred-chart apresentam vantagens em relação a gráficos mais tradicionais. Uma dessas vantagens é a flexibilidade de aplicações, uma vez que para implementar os dois gráficos não ficamos presos a pressupostos estocásticos quanto ao modelo de regressão (isto é, não se faz necessário conhecer a função que relaciona a variável resposta às variáveis explicativas), o que possibilita a aplicação dessas propostas a qualquer tipo de modelo preditivo, não somente aos 16 neste trabalho mencionados. Outra vantagem é a capacidade de elaborar controle de previsões futuras, uma vez que tanto o class-chart quanto o pred-chart mo-

nitoram resultados preditivos. No caso especificamente do class-chart, além das vantagens mencionadas, essa ferramenta possibilita a análise individualizada de múltiplas categorias de interesse ao mesmo tempo, o que facilita o entendimento visual para usuários que não são necessariamente conhecedores da área de Estatística. Essa visualização simultânea e independente de cada categoria dentro de cada amostra do processo, possibilitaria, por exemplo, a criação de um painel iterativo, ou até mesmo a implementação em uma plataforma de BI (do inglês Business Intelligence). Posto isto, em suma, pode-se concluir que a principal atribuição dos gráficos propostos é viabilizar a identificação ágil de problemas nas amostras ou observações, respaldando o usuário quanto a assertividade das tomadas de decisões dos processos. Além disso, as propostas desenvolvidas neste trabalho deixam como contribuição à academia, duas novas ferramentas de controle, elaboradas com conceitos recentes como AM, porém com forte viés matemático, o que auxilia na desmistificação da desassociação da Matemática aos novos implementos tecnológicos.

Contudo, para trabalhos futuros, também é interessante tratar de situações não abordadas neste trabalho, como a dependência temporal entre as variáveis, uma vez que tanto os procedimentos tradicionais quanto os propostos neste trabalho não lidam com a questão de amostras autocorrelacionadas. Outra desvantagem interessante de ser estudada futuramente é quanto ao custo computacional. Testar a implementação de algoritmos genéticos no processo de calibração dos parâmetros do gráfico, pode deixar os algoritmos propostos mais escaláveis para situações de dados massivos, reduzindo o número de iterações necessárias para atender à política de qualidade especificada [19,20].

Por fim, como proposta de trabalhos futuros, existe toda a teoria de Capabilidade (ou Capacidade) do Processo, envolvendo Índices de Capabilidade e conceitos de Processos Não-Centrados, bem como a possibilidade de expansão dos algoritmos para análises multivariadas. Além da aplicação dos mesmos gráficos aqui propostos, porém utilizando diferentes modelos de regressão: mais robustos ou até mesmo combinação de modelos. Tanto os estudos de simulação quanto os estudos de casos reais fornecidos para ilustrar as abordagens propostas e demonstrar sua validade,

foram conduzidos utilizando o software R versão 3.6.1 (R Core Team, 2019) [21]. Os códigos R para aplicação das novas ferramentas CEP podem ser encontrados no GitHub de um dos autores (ver: <https://github.com/llboaventura/class-chart>; <https://github.com/llboaventura/pred-chart>). Para maiores detalhes sobre os procedimentos propostos, poderão ser consultados os artigos publicados recentemente sobre os dois gráficos de controle, separadamente [19,20].

## Agradecimentos

Os autores agradecem à Comissão Organizadora do XXX Congresso de Matemática Capricornio (COMCA 2022) pelo convite e oportunidade de participar do evento, apresentando o presente trabalho (dissertação de mestrado condensada em dois artigos científicos já publicados) sob o título “On flexible Statistical Process Control with Artificial Intelligence: Classification and prediction control charts”, em palestra proferida pelo autor Paulo Henrique Ferreira da Silva, na Sessão Convidada “Estatística”.

## Referências

- [1] Mandel, B. J. (1969). The regression control chart. *Journal of Quality Technology*, 1(1), 1-9.
- [2] Haworth, D. A. (1996). Regression control charts to manage software maintenance. *Journal of Software Maintenance: Research and Practice*, 8(1), 35-48.
- [3] Vining, G. (2009). Technical Advice: Phase I and phase II control charts. *Quality Engineering*, 21(4), 478-479.
- [4] Montgomery, D. C. (2013). *Introduction to statistical quality control*. John Wiley & Sons.
- [5] Kanagawa, A., Tamaki, F., & Ohta, H. (1993). Control charts for process average and variability based on linguistic data. *The international journal of production research*, 31(4), 913-922.
- [6] Franceschini, F., & Romano, D. (1999). Control chart for linguistic variables: a method based on the use of linguistic quanti-

- fiers. *International journal of production research*, 37(16), 3791-3801.
- [7] Balestrassi, P. P. (2000). Identificação de padrões em gráficos de controle estatístico de processos, em tempo real, utilizando séries temporais e redes neurais artificiais. Tese (Doutorado em Engenharia de Produção), Universidade Federal de Santa Catarina. 216 f.
- [8] Gülbay, M., Kahraman, C., & Ruan, D. (2004).  $\alpha$ -Cut fuzzy control charts for linguistic data. *International journal of intelligent systems*, 19(12), 1173-1195.
- [9] Moreira Junior, F. J. (2005). Proposta de um método para o controle estatístico de processo para observações autocorrelacionadas [dissertação]. *Porto Alegre (RS): Universidade Federal do Rio Grande do Sul*.
- [10] El-Midany, T. T., El-Baz, M. A., & Abd-Elwahed, M. S. (2010). A proposed framework for control chart pattern recognition in multivariate process using artificial neural networks. *Expert Systems with Applications*, 37(2), 1035-1042.
- [11] Cheng, H. P., & Cheng, C. S. (2007). A support vector machine for recognizing control chart patterns in multivariate processes. In *Proceedings the 5th Asian Quality Congress* (pp. 17-18).
- [12] Mostajeran, A., Iranpanah, N., & Noorossana, R. (2016). A new bootstrap based algorithm for Hotelling's T2 multivariate control chart. *Journal of Sciences, Islamic Republic of Iran*, 27(3), 269-278.
- [13] Wang, D., & Hryniewicz, O. (2015). A fuzzy nonparametric Shewhart chart based on the bootstrap approach. *International Journal of Applied Mathematics and Computer Science*, 25(2), 389-401.
- [14] Cobb, B. R., & Li, L. (2019). Bayesian network model for quality control with categorical attribute data. *Applied Soft Computing*, 84, 105746.
- [15] Sena, J.G.S.; Ferreira, P.H.; Fiaccone, R.L. (2022). Statistical Process Control as a Tool to Control and Prevent Malaria Epidemics in the Legal Amazon Region. *Brazilian Journal of Biometrics*, 40(1)-
- [16] Woodall, W. H., & Montgomery, D. C. (1999). Research issues and ideas in statistical process control. *Journal of Quality Technology*, 31(4), 376-386.
- [17] Amirzadeh, V., Mashinchi, M., & Yaghoobi, M. A. (2008). Construction of control charts using fuzzy multinomial quality. *Journal of mathematics and statistics*, 4(1), 26.
- [18] Hastie, T., Tibshirani, R., & Friedman, J. (2009). Linear methods for regression. *The elements of statistical learning: Data mining, inference, and prediction*, 43-99.
- [19] Boaventura, L. L., Ferreira, P. H., & Fiaccone, R. L. (2022). On flexible statistical process control with artificial intelligence: classification control charts. *Expert Systems with Applications*, 194, 116492.
- [20] Boaventura, L. L., Fiaccone, R. L., & Ferreira, P. H. (2022). Prediction Control Charts: A New and Flexible Artificial Intelligence-Based Statistical Process Control Approach. *Annals of Data Science*, 1-34.
- [21] R Core Team (2019). R: A Language and Environment for Statistical Computing (Version 3.6.1). R Foundation for Statistical Computing, Vienna, Austria. R Foundation for Statistical Computing. <https://www.R-project.org/>