

REVISÃO DA LITERATURA SOBRE CARTAS DE CONTROLE MULTIVARIADAS NÃO PARAMÉTRICAS

TAMIRES FERNANDA BARBOSA NUNES¹; ARIANE FERREIRA PORTO ROSA²

¹Universidade Federal de Pelotas – tamiresfnunes@gmail.com

²Universidade Federal de Pelotas – afprosa61@gmail.com

1. INTRODUÇÃO

As cartas de controle multivariadas não paramétricas (CCMNP) abrangem técnicas que não pressupõem que determinado conjunto de dados seguem alguma distribuição particular. Segundo HE e WANG (2007) as CCMNP são construídas a partir dos dados de aprendizagem do processo, sem nenhuma suposição sobre a distribuição das variáveis. Para VERDIER e ROSA (2011) a implementação de cartas de controle multivariadas paramétricas supõem que determinado conjunto de observações sigam uma distribuição de probabilidade normal multidimensional.

Em processos industriais a hipótese de normalidade geralmente pode ser de difícil verificação, tornando ineficiente as cartas de controle multivariadas paramétricas, o que torna as CCMNP uma alternativa para o controle e monitoramento do desempenho destes processos. A aplicação de procedimentos não paramétricos ou livres de distribuição possuem a vantagem de não requerer nenhuma hipótese sobre a distribuição dos dados.

As CCMNP são úteis quando não queremos que a função de probabilidade seja modelada parametricamente, seja em função da falta de conhecimento de sua verdadeira distribuição de probabilidade ou para que seja verificada o modelo de paramétrico adotado.

O presente trabalho propõe uma revisão literária, visando analisar a disseminação e expansão das CCMNP. Nos próximos tópicos do artigo serão analisadas as publicações recentes sobre CCMNP baseadas em K-vizinho mais próximo, Cúpulas e Kernel.

1.1. REGRA K – VIZINHO MAIS PRÓXIMO

A regra k vizinhos mais próximos (KNN, do inglês, *K Nearest Neighbors Rule*) é um método de classificação supervisionada, no qual um objeto novo é classificado através da análise das distâncias de amostras de treinamento de vizinhos mais próximos num determinado espaço de características. Assim amostras sem rótulos são classificadas com base em suas semelhanças com as amostras de um conjunto de treinamento. A métrica mais usual utilizada na aplicação da regra é a distância Euclidiana, entretanto se faz necessária uma reflexão sobre a escolha de uma métrica de distância (D) a ser escolhida de forma apropriada para os dados.

O método de classificação de amostras KNN tem sido utilizado em aplicações na área de mineração de dados, reconhecimento de padrões, no processamento de imagens, entre outros. HOLSBACH et al. (2014) a regra aplicada a estudos de diagnóstico precoce de câncer de mama. No âmbito industrial a regra KNN adaptada para detectar de falha nos processos. HE e WANG (2007), VERDIER e ROSA (2011), e LI e ZHANG (2014) utilizam a abordagem na fabricação de semicondutores.

MAGNUSSEN et al. (2010) e McROBERTS (2012) utilizam a regra em inventários florestais. GHARUM et al. (2015) utiliza a abordagem na previsão de rendimento de água após um incêndio florestal. TOWLER et al. (2009) aplica a regra KNN para contornar problemas ocasionados pela variabilidade na qualidade da água.

1.2. CÓPULAS

Métodos baseados em Cópulas permitem a modelagem da dependência dos dados separadamente, em densidades marginais e em dependência conjunta. Um método não paramétrico de estimativa cópula modela a estrutura de dependência, para encontrar a forma cópula que irá convergir no sentido probabilístico formal à estrutura subjacente de dependência. Segundo SILVA JUNIOR (2012) a função de distribuição multivariada cópulas apresenta informações sobre a distribuição conjuntas não presentes em distribuições marginais.

Para VERDIER (2013) a modelagem de cópulas é uma ferramenta estatística de análise multivariável muito eficiente, com diversas aplicações em finanças, ciências, atuarial, entre outras. BOCKER (2008) utiliza cópulas para modelar o risco operacional enfrentado pelas instituições financeiras. DANAHER e SMITH (2011) utilizam os modelos de cópulas de probabilidade multivariada aplicado à literatura de marketing. CHEN et al. (2017) aplicam cópulas como suporte técnico para gestão e controle de tráfego. GRIMALDI et al. (2016) utilizam a abordagem cópulas sobre estruturas de dependência das respostas hidrológicas, identificando semelhanças sob diferentes bacias e cenários.

1.3 KERNEL

Método para estimação da função densidade de probabilidades de uma variável aleatória, no qual as densidades são estimadas através de um núcleo, denominado Kernel. Segundo WANDERLEY (2013) os estimadores kernel tem o intuito de estimar densidades baseadas em informações locais, em vez de realizar a estimação de parâmetros globais para determinado modelo de dados.

KHEDIRI et al. (2011) utilizam a função kernel para o controle *on line* de processos multivariáveis não lineares não estacionários. THELJANI et al. (2015) aplicam um algoritmo kernel para o monitoramento de processos químicos. MARCONDES FILHO et al. (2011) realiza o monitoramento de processos não lineares em batelada a partir da abordagem kernel. HOFFMANN (2006) utiliza a abordagem em manuscritos e mamacitologia do câncer. JONDIKO et al. (2016) aplica a abordagem kernel na estimativa de previsão de qualidade da fabricação de tortas de trigo.

2. METODOLOGIA

Uma pesquisa qualitativa foi desenvolvida, na qual serão observadas as tendências dos avanços científicos investigados durante a análise bibliográfica. Sendo uma pesquisa de natureza básica, focada na geração de conhecimentos atuais para o avanço da ciência sem aplicação prática. “Pesquisa básica objetiva gerar conhecimentos novos úteis para o avanço da ciência sem aplicação prática prevista. Envolve verdades e interesses universais[...].” (SILVA, 2005, p.20)

O desenvolvimento da pesquisa exposta se deu através de uma revisão bibliográfica, realizada a partir de pesquisas em livros e publicações existentes em mídias digitais, com o intuito de fundamentar e tornar explícito a proposta de

pesquisa. A revisão literária considerou as publicações realizadas entre 2006 e 2016, pesquisadas nas principais bases de dados de análise científica.

3. RESULTADOS E DISCUSSÃO

A crescente competitividade empresarial vem moldando a indústria e pressionando as empresas a investirem em diferenciais capazes de torna-las competitiva diante da concorrência. Esta demanda influência diretamente nos processos envolvidos no desenvolvimento de produtos e serviços.

As CCMNP se tornam uma alternativa para os processos produtivos atuais, que apresentam um alto grau de complexidade devido a evolução tecnológica, desenvolvimento industrial, automatização dos processos e características dos dados presentes nos processos, como dependências dos dados. Sendo capazes de aumentar o desempenho dos processos, aumento sua capacidade produtiva e sua confiabilidade, reduzindo os custos, o desperdício e o retrabalho. Tornando o planejamento e controle de produção mais efetivos e eficazes, impactando diretamente na qualidade da entrega final do produtos e serviços, na satisfação dos clientes e na lucratividade da empresa.

A revisão literária das CCMNP e suas implementações recentes propõe enfatizar a necessidade do estudo do estado da arte atualizado sobre o tema. A partir do estado da arte atual sobre o tema pesquisa proposto torna-se possível o surgimento de novas ideias orientadoras de novas pesquisas sobre o mesmo. A avaliação da produção científica nos dá um embasamento para busca de investimentos financeiros para realização de pesquisas, sendo capaz de aumentar a participação da Ciência na construção do conhecimento econômico, social, político e educacional do país.

4. CONCLUSÕES

O controle estatístico de processos por muito tempo tem sido crucial para manter a qualidade e o alto desempenho dos processos das industriais. O alto grau de complexidade dos processos tem se tornado cada vez maior, suas variáveis, suas limitações e os procedimentos envolvidos tem se tornado-os cada vez mais robustos. Assim, é cada vez mais necessário a utilização e implementação de métodos estatísticos que melhor se ajustem aos processos da empresa. Em vista disso, analisar a disseminação de técnicas estatísticas de controle de processos e sua expansão nos mais variados processos e segmentos, se torna um estudo relevante para compreender as demandas atuais dos processos industriais como um todo. Técnicas que não pressupõem que determinado conjunto de dados siga determinada distribuição, tornam-se mais abrangentes a vários processos, dos mais diversos segmentos, se tornando alternativas eficazes para manter um alto grau de desempenho e qualidade nos processos.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- BOCKER, C. K. K. **Modeling and measuring multivariate operational risk with Levy copulas.** Journal of Operational Risk, 2008.
- CHEN, M. et al. **A copula-based approach for estimating the travel time reliability of urban arterial.** Transportation Research Part C, v.28, p. 1-23, 2017.
- DANAHER, P. J.; SMITH, M. S. **Modeling multivariate distributions using copulas: Applications in marketing.** Marketing Science Celabrating 30 Years, v. 30, n. 1, p. 4-21, 2011.

- GHARUN, M.; AZMI, M.; ADAMS, M. A. **Short-term forecasting of water yield from forested catchments after bushfire: A case study from Southeast Australia.** *Water*, v. 7, 2015.
- GRIMALDI, S. et al. **Catchment compatibility via copulas: A non-parametric study of the dependence structures of hydrological responses.** *Advances in Water Resources*, v. 90, p. 116-133, 2016.
- HE, Q.P.; WANG, J. **Fault Detection Using the k-Nearest Neighbor Rule for Semiconductor Manufacturing Processes.** *IEEE Transactions on semiconductor manufacturing*, vol. 20, n. 4, 2007.
- HOFFMANN, H. **Kernel PCA for novelty detection.** *Pattern Recognition*, v. 40, p. 863-874, 2006.
- HOLSBACH, N.; FOGLIATTO, F. S.; ANZENELLO, M. J. **Método de mineração de dados para identificação de câncer de mama baseado na seleção de variáveis.** *Ciência saúde coletiva*, v. 19, n. 4, 2014.
- JONDIKO, T. O. et al. **Prediction of wheat tortilla quality using multivariate modeling of kernel flour, and dough properties.** *Innovative food science and emerging technologies*, v. 34, p.9-15, 2016.
- KHEDIRI, I. B.; LIMAN, M.; WEIHS, C. **Variable window adaptative kernel principal componente analysis for nonlinear nonstationary process monitoring.** *Computers & Industrial Engineering*, v. 61, p. 437-446, 2011.
- MAGNUSEN, S.; TOMPOPO, E.; MCROBERTS, R. E. **A model-assisted k-nearest neighbour approach to remove extrapolation bias.** *Scandinavian Journal of Forest Research*, v. 25, p. 174-184, 2010.
- MCROBERTS, R. E. **Estimating forest attribute parameters for small áreas using nearest neighbors techniques.** *Forest Ecology and Management*, v. 272, p. 3-12, 2012.
- MARCONDES FILHO, D.; FOGLIATTO, F. S.; OLIVEIRA, L. P. L. **Gráficos de controle multivariados para monitoramento de processos não lineares em bateladas.** *Produção*, v. 21, n. 1, p. 132-148, 2011.
- TOWLER, E. et al. **Simulating Ensembles of source water quality using a k-nearest neighbor resampling approach.** *Environmental Science & Technology*, v.43, p. 1407-1411, 2009.
- SILVA, E. L.; MENEZES E. M. **Metodologia da pesquisa e elaboração de dissertação.** Florianópolis: Universidade Federal de Santa Catarina – UFSC, 4º edição, 2005.
- SILVA JUNIOR, J. C. A. **Contágio entre mercados financeiros: Uma análise via cónulas não paramétricas.** Dissertação. Programa de Pós- Graduação em Ciências Econômicas – UFRGS, 2012.
- THELJANI, F. **A new kernel-based classification algorithm for systems monitoring: Comparison with statistical process control methods.** *Arabian Journal for Science and Engineering*, v. 40, p. 645-658, 2015.
- VERDIER, G. **Application of copulas to multivariate control charts.** *Journal of statistical planning and inference*, v. 143, p. 2152-2159, 2013.
- VERDIER, G.; ROSA, A.F.P. **Adaptive Mahalanobis Distance and k-Nearest Neighbor Rule for Fault Detection in Semiconductor Manufacturing.** *IEEE Transactions on Semiconductor Manufacturing*, v.24, n.1, p.59-68, 2011.
- WANDERLEY, M. F. B. **Estudos de estimação de densidade por kernel: Métodos de seleção de características e estimação do parâmetro suavizador.** Tese. Programa de Pós-Graduação em Engenharia Elétrica – Universidade Federal de Minas Gerais, 2013.