

Universidade de Pernambuco

Programa de Pós-Graduação em Engenharia da Computação (PPGEC)

Proposta de Projeto de Doutorado

Área: Computação Inteligente

Título: Detecção de Anomalia em Dados Multidimensionais.

Orientador – Alexandre Magno Andrade Maciel (amam@ecomp.poli.br)

Descrição

Há séculos buscamos meios para identificar padrões estranhos dos dados. As discrepâncias podem surgir por diversos fatores: falhas mecânicas, mudanças no comportamento do sistema, comportamento fraudulento, erro humano, erro de instrumento ou simplesmente devido a desvios naturais nas populações. Sua detecção pode identificar falhas e fraudes do sistema antes que elas aumentem com consequências potencialmente catastróficas [1].

Um outlier pode ser definido como um padrão que não está em conformidade com o comportamento normal esperado. Mas vários fatores tornam essa abordagem aparentemente simples em algo bem mais complexo. Definir uma região que englobe os comportamentos normais é muito difícil. A fronteira entre o comportamento normal e o periférico é confusa. A noção exata de um outlier é diferente para cada conjunto de dados. A disponibilidade de dados para treinamento e validação geralmente é um problema. Em vários casos os valores extremos são resultado de ações intencionais, as fraudes [2]. Frequentemente, os dados contêm ruído semelhante aos valores discrepantes.

Entre as oportunidades de pesquisa apontados pela literatura estão o uso de aprendizagem de máquina [3] para obter *insights*, bem como *analytics* para comparar diferentes abordagens e avaliar o impacto de inovações. Além disso, diversos trabalhos apresentam soluções envolvendo os mais diversos recursos para identificação de outliers.

A presente proposta objetiva usar técnicas contemporâneas para detecção de valores extremos. Deverá ser realizada uma visão abrangente da literatura atual de detecção de outlier seguida de uma investigação da precisão dessas técnicas na eficácia e na eficiência de detecção em casos reais. Dessa forma, espera-se contribuir para a proposição de possíveis novas técnicas e/ou metodologias para solucionar um grave problema: a dificuldade de encontrar anomalias em dados multidimensionais [4].

Referências Bibliográficas

- [1] Varun Chandola, Arindam Banerjee, Arindam Banerjee, Vipin Kumar. Anomaly detection: A survey. ACM Computing Surveys (CSUR) Volume 41, Issue 3. 2009.
- [2] Yufeng Kou, Chang-Tien, S. Sirwongwattana, Yo-Ping Huang. Survey of fraud detection techniques. IEEE International Conference on Networking, Sensing and Control, 2004.
- [3] Ariel Faigon, Krishna Narayanaswamy, Jeevan TAMBULURI, Ravi Ithal, Steve Malmskog, Abhay Kulkarni. Machine learning based anomaly detection. 2019.
- [4] D. Dasgupta, N.S. Majumdar. Anomaly detection in multidimensional data using negative selection algorithm. Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No.02TH8600)