Active Learning in Process Monitoring Using Streaming Data

C. Capezza[1], ***A. Lepore***[1], ***K. Paynabar***[2]

[1]Department of Industrial Engineering, University of Naples Federico II, Italy
[2]School of Industrial and Systems Engineering, Georgia Tech, Georgia, United States

Industrial systems increasingly rely on continuous data streams to monitor process conditions and ensure quality control. While supervised statistical process monitoring techniques can detect deviations from normal operating conditions, their effectiveness depends on labeled data, which are often costly or infeasible to acquire in real time. This work introduces a novel active learning framework for sequential decision-making in labeling under budget constraints. Leveraging partially hidden Markov models, the method captures temporal dependencies and integrates partially labeled data to support dynamic classification of process states. A dual-objective strategy is developed to guide real-time labeling decisions, balancing exploitation, i.e., improving discrimination between known states, and exploration, i.e., discovering previously unobserved out-of-control modes. Validation through simulations and an industrial case study on resistance spot welding in automotive manufacturing demonstrates improved classification accuracy and efficient label usage compared to traditional monitoring approaches.

Keywords: Statistical Process Monitoring, Hidden Markov Model, Sequential Data Analysis, Imbalanced Classification

Acknowledgements
The research activity of C. Capezza and A. Lepore was supported by Piano Nazionale di Ripresa e Resilienza (PNRR) - Missione 5 Componente 2, Investimento 1.3-D.D. 1551.11-10-2022, PE00000004 within the Extended Partnership MICS (Made in Italy - Circular and Sustainable). This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.