Improving the Prediction Accuracy of Data-Driven Fault Diagnosis for HVAC Systems by Applying the Synthetic Minority Oversampling Technique

Shahrad Shakerian,[1] Houtan Jebelli,[2] and William E. Sitzabee[3]

[1]Ph.D. Candidate, Department of Architectural Engineering, Pennsylvania State University, University Park, PA, USA, E-mail address: sqs6690@psu.edu

[2]Assistant Professor, Department of Architectural Engineering, Pennsylvania State University, University Park, PA, USA, E-mail address: hjebelli@psu.edu

[3]Vice President and Chief Facilities Officer, Professor of Architectural Engineering, Pennsylvania State University, University Park, PA, USA, E-mail address: wes25@psu.edu

ABSTRACT

Faulty operation of heating, ventilation, and air conditioning (HVAC) systems in a building can lead to thermal discomfort, wasted energy, and shorter equipment life. Early fault diagnosis in HVAC systems is critical to maintaining indoor environmental comfort, saving energy, and preventing further deterioration of the system. Data-driven predictive models have emerged as a popular approach to fault diagnosis of HVAC systems. For an HVAC system, failures are a small probability event. Therefore, the prediction accuracy of these data-driven approaches is affected by imbalanced datasets. The normal data far outweigh the fault data. Predictive models trained with these imbalanced datasets will perform poorly in fault diagnosis of an HVAC system. To address this limitation, this study aims to examine the potential of the Synthetic Minority Oversampling Technique (SMOTE) for sampling fault data points to improve the accuracy of the fault predictive model. To that end, six months of operational data (for example, humidityand temperature) were collected from embedded sensors in the HVAC system. A SMOTE algorithm was applied to increase the ratio of failure to normal data from 0.02 to 0.3. Both the original and improved datasets were used to train the fault predictive model based on different supervised learning algorithms. Results indicated that datasets improved through the SMOTE algorithm increased the accuracy of the predictive model by about 20%. This improvement can lay the groundwork for increasingly proactive facility maintenance.