A combined model-based and data-driven approach for monitoring smart buildings

Hamed Khorasgani and Gautam Biswas

Institute of Software Integrated Systems & Dept. of EECS
Vanderbilt University, USA

Sep 26, 2017

28th Edition of the International Workshop on Principles of Diagnosis (DX)
Presentation Overview

- What are smart buildings?
- Motivation & Goals of this work
- Research Challenges
- Model-based residual analysis for FDI in smart buildings
- Data-driven feature extraction
- A combined approach for FDI in smart buildings
- Conclusions
Smart buildings

- **Smart buildings**: Use IOT technologies to monitor and maintain building performance
  - embedded sensors + interconnected devices + ability to store analyze, and exchange data
  - Key question: how to perform analysis and support decision making?

- **Example**: Outdoor air unit (OAU)
  - Components
    - Exhaust fan
    - Outdoor fan
  - 20 sensors measure:
    - Continuous variables:
      - static pressures in the fans, fan rotational speeds, fan airflows
    - Discrete events:
      - fire alarm, fan filter status, fan status (on/off)
Motivation & Goals

• **Motivation: monitoring of smart buildings to**
  – Avoid unnecessary wastage of resources
  – Avoid discomfort for residents
  – Prevent extended downtimes

• **Goal: Design a Diagnosis Approach that**
  – Can operate in an uncertain environment
  – Does not require complete knowledge of the system
  – Updates as the system operates
  – Generates correct results

  – In other work, we have also developed methods for data-driven energy monitoring
Research Challenges

• **It is not feasible to generate an accurate and complete model for smart buildings**
  
  – Especially difficult because of the highly precise and accurate spatio-temporal models that have to be created
    
    • May require millions of dollars & many years just to build models
  
  – Can we do it for components and subsystems?
    
    • Outdoor air unit (OAU)
      
      – Relationship between a fan’s static pressure and airflow is nonlinear and a function of the fan’s rotational speed.
      
      – The performance of the exhaust fan and the output fan are not independent but the dependency is not modeled.
      
      – Unknown parameters such as wind speed, and the air filter’s resistance affect the model.

• **May not have training data for all the operation modes and fault modes**

9/26/2017  

DX-17: Brescia, Italy
Solution Approach

- **Combine model & data driven approaches**
  - Models → Models + Data
    - Use models when sufficiently accurate models are available
    - Enhance models with operational data when required
  - Works for engineered + well-circumscribed subsystems
  - **What about subsystems with complex spatio-temporal relations?**
    - Accurate models based on complex nonlinear (and often empirical) flow relations
    - Finite element models
  - Resort to pure data driven models
  - Decision Trees, Regression Trees, Naïve Bayes, Support Vector Machines, Neural Networks, etc. for **supervised analysis**
  - Semi supervised and unsupervised **anomaly detection**,
Model-based Fault Detection and Isolation

- **Model-based Approaches:**
  - Use a physics-based model that defines nominal/faulty behavior of a dynamic system to detect faulty behaviors.

**Residual**: A fault indicator, based on a deviation between measurements and model-equation based computations.

**Hypothesis test**: determines when change in a residual values are statistically significant.
Model-based Fault Detection and Isolation in OAU

- **Faults**
  - Only one fan is operating (in normal situation they are both on or off)
  - Exhaust fan or outdoor fan filters are dirtyBlocked

- **Diagnoser design:**
  - The complete model was not available
    - Used laws of physics to derive relationships between fan speed, static pressure, and airflow
    - Developed a maximum likelihood estimator (MLE) to estimate the parameters
    - Analytical redundancy relationship (ARR) approach to generate the residuals
    - Z-test [Biswas et al., 2003] as the hypothesis test
• Physical laws to derive relations between exhaust fan, outside fan speed, static pressure and airflow

\[
P_2 = P_1 \left( \frac{D_2}{D_1} \right)^2 \left( \frac{N_2}{N_1} \right)^2 \left( \frac{\rho_2}{\rho_1} \right); \quad Q_2 = Q_1 \left( \frac{D_2}{D_1} \right)^3 \frac{N_2}{N_1}
\]

\[
Q_i \rightarrow \text{airflow}; \quad D_i \rightarrow \text{diameter}; \quad N_i \rightarrow \text{rotational speed}; \\
P_i \rightarrow \text{static pressure}; \quad \rho_i \rightarrow \text{air density} \quad \text{– for fan } i
\]

Equations + 6 sensors used to derive 4 residuals
Total Number of Measurements

- **Measurements:** 6 continuous + 14 binary-valued time series waveforms

- **Sampling rate** – 6 samples/hour

- **Used about 3 months of data for our study (11,193 samples)**
  - Data had missing values – Removed them during preprocessing – 10,316 samples
Model-based Fault Detection and Isolation in Output Air Unit

Diagnosis reference model using

We use a data mining approach to extract additional features from the data in order to improve diagnosis performance.
Unsupervised Data-driven Feature Extraction

- **Preprocessing**
  - Standardizes the time series variables
- **Clustering**
  - Extracts the clusters in the data set
- **Significant Features**
  - Set of features that best distinguish an anomalous cluster from nominal operations

Operational Data

- Preprocessing
- Clustering

Nominal Groups

Anomalies

Distance Metrics

Mack, et al, DX-16
Biswa, et al., IJPHM 2016
Preprocessing

• **Removes clustering sensitivity to the amplitude of the input signals**
  
  – Approaches which standardize by division by the range of the variable give superior performance in recovering the clusters [Milligan and Cooper, 1988]

  – For each feature F:

\[
F_s = \frac{F - \min(F)}{\max(F) - \min(F)},
\]
Clustering

• **Density-based clustering (DBSCAN)**
  - Can extract clusters with arbitrarily shapes
  - Automatically determines the clusters and the outliers
  - Inputs:
    • MinObj (Minimum objects in each cluster)
      - We consider MinObj = 10.
    • Reachable distance $\epsilon$
      - Sharp change in K-dist plot ($\epsilon = 0.02$)

5-dist plot for the AOU
We apply the `dbscan2` – R-package

Clusters for the OAU system (total training samples: 10,316)

- Nominal group (6299 samples cluster 1)
- Anomalous groups (clusters 2-6)
  - Cluster 2 (852 samples): system is off
  - Cluster 3 (19 samples): transition (off to on)
  - Clusters 4 (2968 samples): outdoor fan filter is dirty
  - Cluster 5 (55 samples): exhaust fan filter is dirty
  - Cluster 6 (59 samples): exhaust fan off/ outdoor fan on
- Outliers (64 samples (cluster 0))
  - We did not analyze outliers

Clusters (1-6) and outliers (0).
**Definition 1** (Significant features). *Significant features are a single feature or a set of features that best distinguish an outlier group from nominal operations of a system.*

- **Selecting the significant features for each cluster:**
  - We use the $k$–nearest objects to define the distance between two clusters.
  - The importance of each feature is the ratio of its distance in two clusters over the overall cluster distances.
  - The subset of features that account for 90% of the distance between two clusters are the significant features.
# The Operating Modes & Their Significant Features

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Detected Anomaly</th>
<th>Mode or Anomaly</th>
<th>Significant Features</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>Normal operation mode</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| 2       |                  | Mode: the system is off |  ● Exhaust fan status  
          |                   |                  |  ● Outdoor fan status |  ● The OAU is off in this mode |
| 3       |                  | Mode: transition |  ● Outdoor fan static pressure  
          |                   |                  |  ● Outdoor fan airflow  
          |                   |                  |  ● Outdoor fan speed command |  ● Low pressure and airflow when the system starts |
| 4       |                  | Fault: the outdoor fan filter is dirty |  ● Outdoor fan filter status |  ● The outdoor fan filter has to be changed. |
| 5       |                  | Fault: the exhaust fan filter is dirty |  ● Exhaust fan filter status |  ● The exhaust fan filter has to be changed. |
| 6       |                  | Fault: only one fan is working |  ● Exhaust fan speed command  
          |                   |                  |  ● Exhaust fan status |  ● Exhaust fan off and outdoor fan on |
Integrated Model + Data driven Fault Diagnosis

- **Model-based diagnosis:**
  - Monitors: outputs of the hypothesis tests

- **Data-driven diagnosis:**
  - Monitors: selected features

- **Integrated approach**
  - Monitors: residuals + significant features

Hybrid diagnosis reference model
We use the training data to learn/update the probability function of the reference model.

- Bayesian Networks (BN)
  - Assumes the monitors are independent
    - A residual can be a function of one or more significant feature
- Tree Augmented Naive Bayesian (TAN)
  - Provides additional links to model the dependencies among the monitors
Diagnosis Results

- Significant Features improve the diagnosis performance in Lentz Public Health Center in Nashville.

<table>
<thead>
<tr>
<th>Diagnosis approach</th>
<th>Accuracy</th>
<th>False positive rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model-based approach</td>
<td>87.1%</td>
<td>12.7%</td>
</tr>
<tr>
<td>Hybrid approach</td>
<td>92.5%</td>
<td>2.5%</td>
</tr>
</tbody>
</table>
Conclusions

• We proposed a combined model-based and data-driven diagnosis method for smart buildings.
• Our approach uses model-based residuals and significant features to detect and isolate faults.
• We developed an unsupervised approach to extract significant features.
  – Can be applied to datasets without labels.
• We used TAN structure to update diagnosis reference model.
• The case study shows the proposed hybrid approach significantly improves the diagnosis accuracy and reduces false positive rate.
Acknowledgement

The authors gratefully acknowledge the support provided by the Department of General Services at the metropolitan government of Nashville in acquiring, analyzing, and interpreting the data.
Thank you
&
Questions (??)