Comparing methods aimed at model based sensor fault detection analysis of smart building systems.

Martijn Bennis, Gudjon Bjarnason, Justin Kharpatoe, Arjan Roos, André Rodrigues Rosa, Jan Dirk Schagen, Rensley Starke, Bob Tuynman, Rowan Voermans, Steven Wetstein,
Faculty ITD, Faculty TIS, The Hague University of Applied Sciences
The Hague, the Netherlands

Abstract

The Hague University in Delft uses an advanced climate control system. All sensors and actuators are monitored and deviations from the default sensor data which are reported daily. The building manager has to combine the information from the sensor data in order to draw the right conclusions. In this paper, we will discuss four different methods and find the best fit based on the question 'How can sensor data be automatically analyzed and presented, so anomalies can be reported in real-time?'. These four methods are: 'Rule Based System', 'Deep Learning', 'Bayesian Belief Network' and 'Clustering method'. First there is a necessity which anomalies are currently taking place. After that there has to be determined how to track down future anomalies. Last is figuring out how found anomalies can be used to track down new ones.

Keywords: Bayesian Belief Networks, Expert System, Deep learning, HVAC system, sensor fault detection

1. Introduction

HVAC systems are getting more advanced, they generate more data which arises opportunity for optimization. This research is based on data from the Living Lab building from The Hague University of Applied Sciences at Delft. The building has a very complex HVAC system which generates data. The data that will be analyzed originates from a range of sensors: CO_2, temperature, occupancy (PIR), light, ventilation rate, ventilation valve position, window open/close. It will be analyzed using different methods, the chosen methods are Rule Based System (RBS), Bayesian model (BBN), Deep Learning and Cluster method. Using data from the past seven years we can analyze how these different methods would operate. The outcome can be compared between methods, comparing these results would extrapolate a preferred method for HVAC systems.

2. Definition of comfort

One of the main goals of the research mentioned above, is to improve the experienced comfort for the visitors of the above mentioned University of Applied Sciences in Delft. The definition of comfort is different for each individual. However, through previously executed research, a statistical definition has been created. Within this research paper the same definition for comfort is implemented. This definition can be found within 'Climate Design: Solutions for Buildings That Can Do More with Less Technology, Hausladen, G. et al., 2005'.

3. Expert System

In previous researches[1] a Rule Based System, hereafter referred to as RBS, has been build to analyze the available data from the climate system of The Hague University of Applied Sciences in Delft (Kortekaas, C., Vuuren, B. van; 2016). By composing variables and rules for a situation in a room we can analyze the data in order to find whether the system is performing within its boundaries. This could be a defect or there could be a different cause. The main focus of the RBS was to function as a validation technique for the lectorate. Using the RBS it is possible to define new methods and evaluate the operation.

For this research on RBS we have performed the following tasks:

- Making the system function properly
- Analyse and visualize the previously obtained results
- Expand the system and increase the accessibility

4. Bayesian Belief Networks

[1] Foutdetectie van sensoren in het klimaatregelsysteem; Vuuren B. van, Kortekaas C.

BBN is mostly used in the medical world (Ashby, B.; 2006). That is why the building is often compared with a diseased patient by the lectureship 'Lectoraat Energie en de Gebouwde Omgeving' ('LEGO'); it has a number of symptoms, and combining them together there is a chance on a disease. This is also applicable on a building; examples of symptoms are: a too high CO_2, too high temperature or too low airflow value. A broken ventilation -, CO_2 sensor or over occupied are examples of results/diseases. By combining the active symptoms and using the theorem of Bayes, the conditional chance of an occurring result can be calculated. By using a chance instead of using a definitive result, multiple possible results can be taken into account.

BBN is in this research used while performing the following tasks:

- Tweaking an C++ application (Wazir, F.; 2017), named smileApp, so it can use the given BBN model from the lectorate LEGO.
- Making a predicting model, which uses a form of supervised learning, in Python.

5. **Deep Learning**

Out of the four chosen methods, 'Deep Learning' was chosen to further answer the research question of Laure Itard, the Lector of lectorate 'Energy and the Built Environment'. The research question is stated thusly: 'Is it possible to develop a fool-proof system that comprehensible for non-experts and able to detect anomalies?' The definition of 'non-expert' within this question is as follows: 'An individual that does not have the same knowledge as someone with professional experience in the field of indoor climate maintenance.' Deep Learning is a type of Machine Learning which is inspired by the human brain. The network can determine what the default sensor values would be in in normal conditions. When enough knowledge is gathered to define normal behavior and sensor value, this model can be used as a baseline. Every value that has a large differentiation from this baseline can be classified as an anomaly.

Certain tasks have been executed in order to further this stated research; these are:

- Research to find out which neural network is suitable for solving the research goal.
- The making of a neural network that is able to detect anomalies within a time-series dataset.

6. **Clustering method**

Clustering is a method which is generally used to group research objects together based on characteristics. This technique can be used to gain insights as to how well the system functions regarding comfort, based on characteristics from the data. Within the stated research an immaculate amount of data has been made available. However, not much is known about the semantics of the stated data. To solve this, literary research has been conducted in order to find a solution to the classification and data clustering problem.

The research showed that the usage of factor analysis is a very useful technique in order to reduce the amount of variables which makes it possible to analyse multiple, even all available, variables at the same time.

To achieve the above, the following tasks have been executed:

- Research into the applicability of 'Factor Analysis' for the available data.
- Literary research into the applicability of different cluster recognition algorithms.

7. **Results**

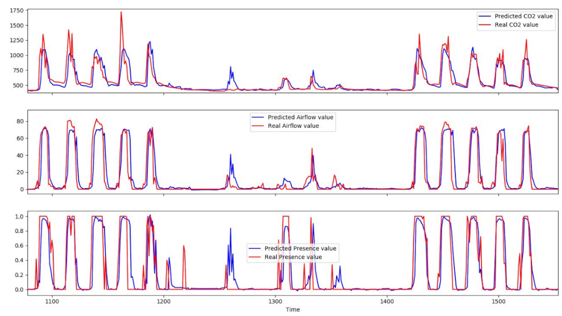
Rule Based System was researched to with the purpose to validate our results. However, due to difficulties in the SAW there was decided to check the results by hand. During a period from the January 1st 2012 till the June 23rd 2016, anomalies where detected in 66 of the 119 rooms. In 64 these 66 cases the CO2 sensor was probably broken.

Because of the trouble with SAW there was decided to write a new application to analyse the available data. The application was written in Angular and a Backend application for the analysis in C#. Unfortunately, due to a lack of time, the application is still under development.

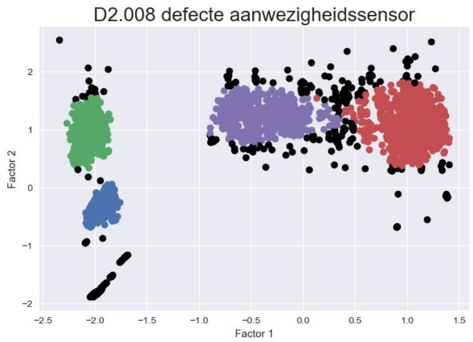
The outcome for BBN in combination with the 'smileApp' is disappointing. After analyzing the data from 28-04-2015 till 01-06-2015, the application predicted a broken damper with a mean of 66 percent over 13 measurements. This was not the expected outcome, because it is known that the PIR sensor was defect during the chosen time period. The weirdest thing is that it never predicted a defective PIR sensor, not even the slightest bit.

BBN in Python uses a form of supervised learning to predict the probability of a broken sensor. The model simplifies these calculations by determining it per attribute, apart from other attributes. This results is a quick and effective method. By multiplying the values of attributes, multiple classes can be defined. When multiple classes are defined, the biggest probability that the values belong to a specific class can be calculated based on the values of multiple attributes.

In Deep Learning, ‘LSTM’ networks would be the most promising to be used for anomaly detection in time series. In figure **X** a LSTM was used to predict a time series of three weeks; a normal school week, followed by a week of holiday and closed by another school week. As shown, the ‘LSTM’ takes a fairly good prediction. However, there are still some particularities in this graph. For example, during the first monday of the holiday the ‘LSTM’ predicts a new school day because of a limited window-size. To solve that problem, the window-size is increased to a week instead of 24 hours. Further optimization can be done by tuning the hyperparameters of the ‘LSTM’ network. The LSTM-model can be used as a baseline. Whenever the real value deviates to much from this baseline, an anomaly is detected



In order to validate the applicability of transforming data by means of Factor analysis to find clusters, a dataset in which an known defective presence sensor has been analysed. The sensor was defect during the time between April 20th and the 1st of June. This period was found by transforming a year of data, including the searched period, into five distinctive “factors”. The found factors were then analysed and grouped into clusters by means of the cluster recognition method known as “HDBscan”. Multiple cluster recognition methods were reviewed whereof the “HDBscan” was the best method to apply. This is concluded because it does not include every analysed data-point it receives, so it does not add random data-points to existing clusters. In the figure below is a ‘scatterplot’ displayed between two of the five factors which resulted in four found clusters, which have been given distinctive colored, and a bunch of points that do not belong to any cluster, given a black color. The blue cluster in this figure is the period in which the presence sensor was defect. This gives proof that transforming and clustering a reasonably sized dataset can be used to detect defective sensors.



8. Conclusion

This research explored in which way The Hague University of applied sciences in Delft could use sensordata to automatically analyse and visualize the data, to find real time anomalies and report them.

The results conclude that the methods ‘RBS’ and ‘Deep Learning’ both displayed useful results, whereas the given ‘BBN’ model from ‘LEGO’ first needs to be revised to showcase similar result. Last was shown that representing and clustering the data may also be a good way to detect anomalies.

The Hague University of applied sciences in Delft also wanted to know what specific anomalies were happening. Out of the data and models we used was the only outcome ‘CO2 sensor is presumably defect’. Through training and expanding the models more anomalies would have been known.

By operating the cluster analyses, two known anomalies have been detected, this showcases the validation and usability of the cluster analyses. The second subquestion; ‘In which way can sensordata be used to find anomalies’. To answer this question four of methods have been researched, these are: ‘RBS, BBN, Deep Learning and cluster analyses. Whereby every method can be used to find anomalies.

For the third and last subquestion which states: ‘In what way can the found anomalies be reported’. In the research it is shown that a web-application have all the functionality to answer this question.

The main question of the research is: How can The Hague University in Delft use it's sensordata to automatically analyse and visualize the data, to find real time anomalies and report them.This question can be solved by using a web-application to show the data, which uses a LSTM method with perhaps a clustering function to find the anomalies. This is because the LSTM and clustering method are the only two methods that can automatically analyse the data without a expert.

9. Discussion

The results of the research into the expansion and improvement of the 'RBS' quickly showed that the user-system 'SAW' (Sensor Application Wrapper) does not meet the requirements. The reason is existing 'bugs' within the system. The biggest disadvantage of this is that the 'RBS' can not be tested on validity. Because 'RBS' is the most used method within the university, it stays the best method to benchmark other methods.

Although it is currently the best method to benchmark with, that does not mean a better method does not exist. Because of this, an attempt has been made to detect the anomalies manually. The attempt was unsuccessful. Therefore, validity has not been achieved.

For both BBN and Deep Learning it was barely possible to validate the received results. This is due to the fact that the used validation system, the RBS, is unable to process the data received from the new database, nor does it allow the new data to be analysed. The new data cannot be analysed because the received application contained certain faults which the research group could not solve. Due to this both models have been validated by a series of known anomalies. Unfortunately, there are but a few known anomalies so this form of validation is not solid enough to guarantee that the results are correct.

For future follow-up research, it is highly advised, to have sufficient knowledge of 'BBN' as well as 'Deep Learning' to be able to fine-tune used methods.

It is also advised to develop a web-application, together with the 'RBS' and in combination with an 'Backend' application for LSTM-models and 'BBN'

Furthermore, it is of utmost importance that a validation method, that is applicable for all detection methods, is found or developed. With such a validation method, detection methods can be validated and benchmarked.

References

[1] Kortekaas, C., & Vuuren, B. van (2016). Foutdetectie van sensoren in het klimaatregelsysteem.

[2] A. Deborah. Bayesian statistics in medicine: a 25 year review (2006) 3589–3631, DOI: 10.1002/sim.2672

[3] F. Wazir. Genie analyzer. (2017)

[4] A. Taal, L. Itard, W. Zeiler, and Y. Zhao. Automatische detectie, diagnose en foutencorrectie. TVVL Magazine 11 (2015) 16-19.

[5] Y. Zhao, F. Xiao, and S. Wang. An intelligent chiller fault detection and diagnosis methodology using Bayesian belief network. Energy and Buildings 57 (2013) 278-288.

[6] A. Carbonari, M. Vaccarini and A. Giretti. Bayesian Networks for Supporting Model Based Predictive Control of Smart Buildings. (2014) in Dynamic Programming and Bayesian Inference, Concepts and Applications, Dr. Mohammad Saber Fallah Nezhad (Ed.), ISBN: 978-953-51-1364-5, InTech, DOI: 10.5772/58470.