Information Visualization

Ronja Grosz, Isabell Jansson and Jonathan Bosson

Abstract— A visualization of the second mini challenge of the 2016 VAST challenge has been made using methods such as k-means, parallel coordinates and variance to find patterns and connections in and between the data.

Index Terms—Multivariate data, Parallel coordinates, k-means, variance.

1 Introduction

The aim of the project was to analyze prox sensor data measured within two weeks in several HVAC (heating, ventilation and air condition) zones in a building. Hazium is a recently discovered possibly dangerous chemical and has been analyzed with the prox sensor data from different perspectives during the project.

2 Background and Related Work

The aim of the project was to analyze prox sensor data measured within two weeks in several HVAC (heating, ventilation and air condition) zones in a building. Hazium is a recently discovered possibly dangerous chemical and has been analyzed with the prox sensor data from different perspectives during the project.

3 Data

The prox sensor data contains information about the hazium concentration in four different HVAC zones in the building, position of the employees registered by both mobile sensors and fixed sensors and HVAC data containing information about the atmosphere in- and outside the building (wind, temperatures etc.). For a detailed list of the HVAC data, see Table 1 in Appendix A. The prox sensors have registered information every fifth minute during a two week period.

Some of the data files contained duplicates of registered measurements. Those were removed in a preprocess through a python script.

4 Method and implementatioin

All necessary data were imported into different collections in a MongoDB database.

The application contains three different views for different analyzing purposes, map, correlation and deviation. A detailed description of each view follows below.

4.1 Map

The map tab is the view which visualizes the GASTech offices. The user can change the view and floor level of the map in order to visualize the employees spatial location in the building.

By clustering with the method k-means by the times a worker has been in a zone with a hazium concentration above zero, as well as recording how hazium concentration level, can the threat level be visualized through color coding. This way can the employees who has been exposed to a dangerous hazium concentration be identified to be treated. The people who has been exposed the most are also shown on a list on the side of the map.

4.2 Correlations

Parallel coordinates are used in order to visualize correlations between the hazium level and HVAC data listed in Appendix A. Only data that was not constant during the time period are visualised. The data can be filtered with a brush in order to highlight a specific region. The axes can be moved to visualize different correlations.

The hazium is measured in four different zones and therefore there are four hazium values for each timestamp. To be able to visualized the hazium level in the plot, the mean hazium level was calculated for every timestamp.

4.3 Deviation

To see if the HVAC and hazium data values had any changes in the data values, variance was used. For each data category each data point is chronologically examined. If the variance of the two previous values are lower than the variance of the examined value and the previous value, the value and the time for this data point is shown in a table associated with the data points day.

The different data categories can be view through choosing the category in a drop down list. Categories from the HVAC data set and the four different hazium zones, where the data is not constant, can be chosen.

5 Results

5.1 Map

The map view can visualize the people at increasing risk from a high hazium exposure as well as give us the names of the people at highest risk. The clustering algorithm used is kmeans, which means that the number of clusters are set by the user. This can result in low quality clusters and hard edges. A high number (more than 6) of clusters ended up giving the best results.

5.2 Correlations

The hazium level is generally low, but this is a relative result and it does not tell if the level is to high. It can be seen that if the Hazium level is high, the Supply Side Inlet Temperature is high. But there are also measurements where a high temperature results in a low Hazium level, the level depends on multiple factors and no result is univocal.

5.3 Deviation

For the deviation the results were different depending on the data category. For categories where the data changed a little for every five minutes, most of the data was still listed in the tables. For categories where the data mostly stayed constant, a lot of data was not shown in the tables.

For example on May 31 2016 between 02:05 - 04:00 the category Supply Side Outlet Temperature has ten data values and the category DELI-FAN Power only has one data value in that interval. This is because the first category has a lot of changes in its values, while the latter category only has a small amount of changes. Therefore the Supply Side Outlet Temperature has more values in its table.

6 Conclusions and Future Work

6.1 Map

The images used for the offices were given in jpeg format. Which means it was difficult to vectorize the image to put coordinate data to it. This meant that the coordinates the points were to use had to be normalized and scaled dynamically with the users resolution, unrelated to

• *Ronja Grosz, Isabell Jansson and Jonathan Bosson is with Linköping University, Sweden, e-mail: rongr946@student.liu.se, isaja187@student.liu.se, jonbo665@student.liu.se.*

the image. A better solution would have been to use an SVG image instead of a raster image to more accurately plot out the points on the map.

6.2 Correlations

The parallel coordinate plot was easy to implement, but the result is complex and it is difficult to reach any conclusion. A 3-Dimensional parallel coordinates plot could possibly be a better solution.

Since no result is univocal, no straightforward conclusion can be drawn. It can be seen that the hazium level is relatively low, but has some peaks. What the peaks depend on has to be further investigated. The results have to be combined, hazium levels in different zones, with the HVAC data. Since there are no guidelines yet about hazium, it is not possible to say whether or not the hazium is a risk for the employees. But it can be seen that a combination of several HVAC factors increases the level.

6.3 Deviation

The deviation was difficult to interpret, since knowledge about the values is needed to know if the changes are abnormal. It also became difficult to interpret changes between days, since you had to check the value in one table and then in another table. Therefore no significant conclusions could be drawn. Presenting a category in a line chart, where a day is represented by a line, would make it easier to spot outliers without prior knowledge of the data.

7 Appendix A

Table 1. HVAC data

HVAC data	**Description**
DELI-FAN Power	Power used by the deli exhaust fan
Drybuld temperature	Drybulb temperature of the outside air
Wind direction	Direction of the wind outside the building
Wind Speed	Speed of the wind outside the building
HEAT Schedule Value	Supply air temperature set point. Air exiting the HVAC system fan is maintained at this temperature during heating operation
Pump Power	Power used by the hot water system pump
Water Heater Setpoint	Water heater set point temperature
Water Heater Gas Rate	Rate at which the water heater burns natural gas
Water Heater Tank Temperature	Temperature of the water inside the hot water heater
Loop Temp Schedule	Temperature set point of the hot water loop. This is the temperature at which hot water is delivered to hot water appliances and fixtures.
Supply Side Inlet Mass Flow Rate	Flow rate of water entering the hot water heater
Supply side Outlet Temperature	Temperature of the water exiting the hot water heater
Supply Side Inlet Temperature	Temperature of the water entering the hot water heater