Learning CO_2 levels as a Partial Indicator of Air Quality in Smart Environments: A Data Mining Task

Seun Deleawe, Dr. Diane Cook, Dr. Brian Lamb, Jim Kusznir

Electrical Engineering and Computer Science

Introduction

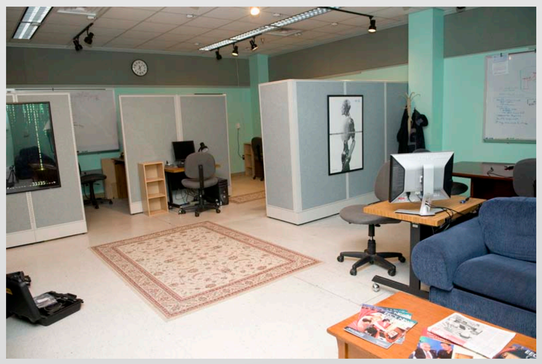
Historically, outdoor air quality has received a disproportionally greater amount of attention than indoor air quality. Nevertheless, a recent study showed that U.S. residents, on average, spend 88% of their day inside buildings, 7% in a vehicle, and only 5% outside. As people continue to spend the majority of their time indoors, air quality continues to have a significant effect on health.

Indoor air quality is often described by the presence or absence of various pollutants. These pollutants include but are not limited to combustion products, volatile organic compounds (VOCs), and biological particles.

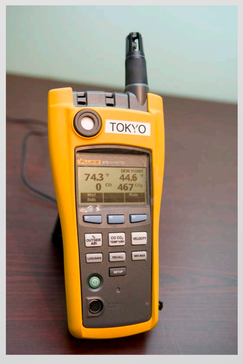
Capable of automating events based available air quality data, smart environments posses the potential to improve the indoor air quality and therefore health of their inhabitants.

Overview

In our study, we monitored Carbon Dioxide (CO_2) levels as a partial indicator of air quality. We used Fluke 975 AirMeters to record CO_2 readings at four different locations codenamed TOKYO, KYOTO ,TULUM, and TOKYO OUT. The former three are smart environments which are equipped with motion sensors. The latter is the outside environments that corresponds with TOKYO. The meters collected data for a period of one week at each location logging one CO_2 reading every one minute for a total of 10,080 readings. We then proceeded to analyze the CO_2 data along with motion sensor data using Weka, a data mining tool.



TOKYO



TOKYO's Fluke 975 AirMeter

activated motion sensor

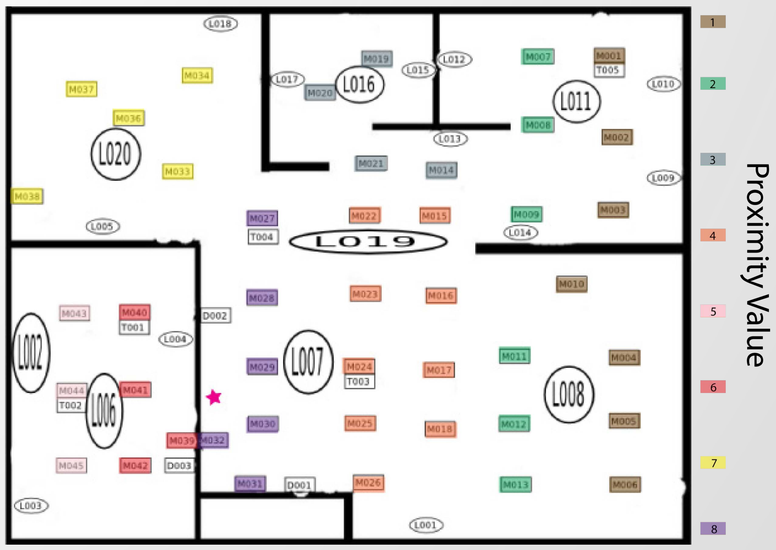


Figure 1: TOKYO weighted sensor map

Data Mining

A significant amount of time was spent preparing data in ARFF format for analysis with Weka. The ARFF files contained a title, attributes, and instances. The attributes had either 2 or 3 labels.

Each sensor was assigned a value from 1-8 based on the proximity of the sensor to the meter (8 being the closest). The weighted values of the sensors activated during a specific time period were then added for a weighted "ON" sensor count.

We created an ARFF file for each smart environment with the attributes: time of day, temperature, number of weighted or un-weighted "ON" sensor events 5x minutes before a given CO_2 reading where $1 \le x \le 6$, and the current CO_2 reading at the smart environment. The ARFF file for TOKYO contained an additional TOKYO OUT attribute.

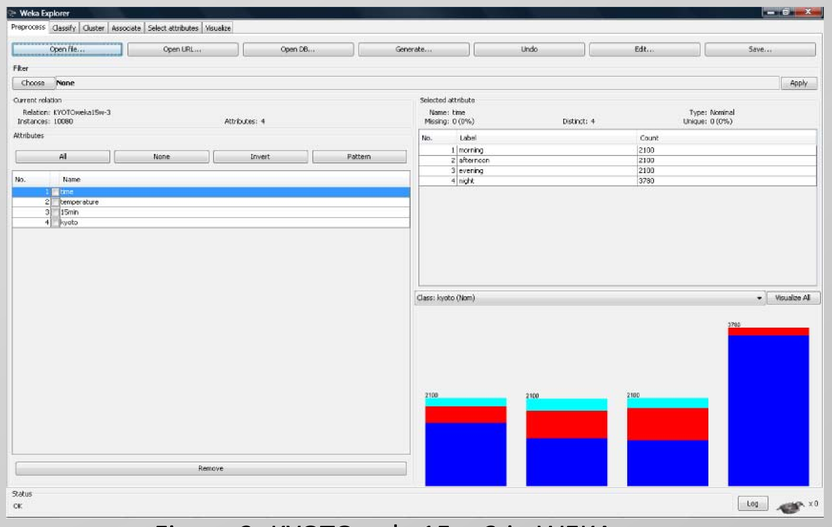


Figure 2: KYOTOweka15w-3 in WEKA

Results

We used both a J48 decision tree and a Naïve Bayes Classifier to analyze how well the CO2 levels in the smart environment could be learned.

	J48	NaiveBayes
TULUMweka5uw-3	99.52%	99.45%
TULUMweka10uw-3	99.52%	99.46%
TULUMweka15uw-3	99.52%	99.42%
TULUMweka20uw-3	99.52%	99.41%
TULUMweka25uw-3	99.52%	99.42%
TULUMweka30uw-3	99.52%	99.31%
TULUMweka5w-3	99.52%	99.45%
TULUMweka10w-3	99.52%	99.46%
TULUMweka15w-3	99.52%	99.42%
TULUMweka20w-3	99.52%	99.41%
TULUMweka25w-3	99.52%	99.52%
TULUMweka30w-3	99.52%	99.31%
MAX	99.52%	99.52%
MIN	99.52%	99.31%

Figure 3: TULUM results for files with 3 labels/attribute

Conclusion

Our results indicate that CO2 levels can be learned with a reasonable amount of accuracy and therefore used as a partial indicator of air quality in smart environments. Generally, files with 2 labels per attribute were learned with better accuracies than files with 3 labels per attribute.

Future improvements upon these results may require more accurate estimates of the proximity of sensors to the meter, better assignment of labels, and more accurate measurement of CO2 levels.

In comparison to the results for TOKYO and KYOTO, the results for TULUM files with 3 attributes were outstandingly exceptional. It would be interesting to investigate what caused this.

Acknowledgement & Citing

Jones, A. P. "Indoor air quality and health." Atmospheric Environment 33 (1999): 4535-564.

•A special thanks Dr. Cook, Brian Lamb, Jim Kusznir for their continued guidance and support

•Thanks to the EECS and LAR departments for permitting the use of their facilities

•Thanks to the National Science Foundation for their support of the REU program under grant number IIS-0647705.