

Mustafa Khalid Masood, Vishal Singh, N. Asokan

Occupant identity leakage from CO₂ sensors

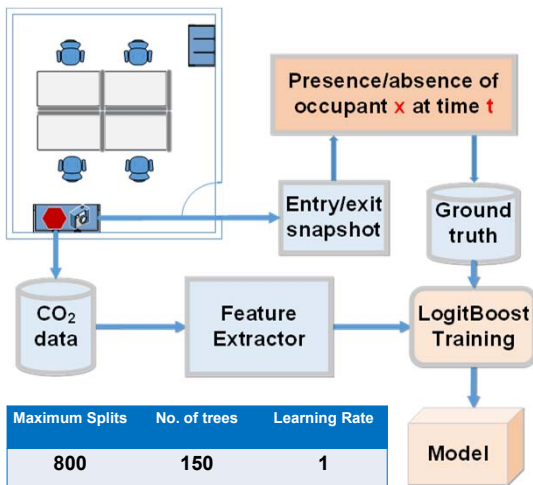


Motivation

- CO₂ data collected in buildings is considered **nonintrusive** to occupant privacy
- However, CO₂ generation rates **vary across individuals** based on height, weight, metabolic rate etc.
- An adversary could use this to **infer occupant identities** from CO₂ data

Methodology

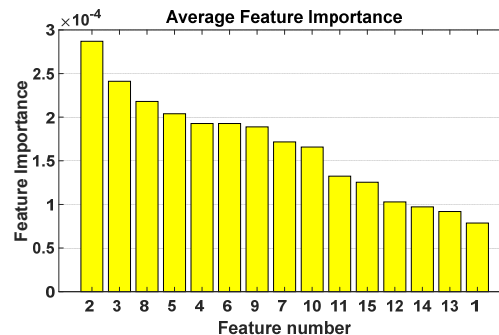
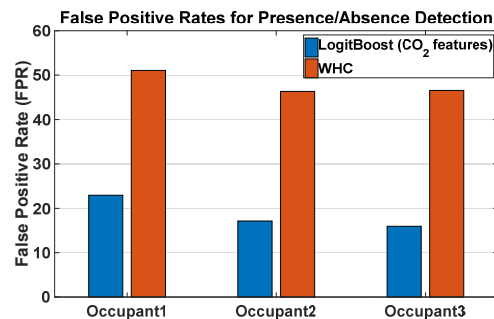
- CO₂ data collected for 21 days in a 4-person office
- Extracted features used to train a LogitBoost classifier
- Ground truth: presence/absence of occupant x ($x = 1,2,3$)



Feature #	Feature type	Description
1	Day of the week	Index for Mon-Fri
2	Time stamp	Time of sensor reading
3-9	Shifted differences	$CO_2(n) - CO_2(n - d)$
10-15	Moving averages	$\sum_{j=n-(m-1)}^n CO_2(j)/m$

Results

- 21-fold cross validation (1 fold=1 day, 6am to 9pm)
- Compared with a Working Hours Classifier (WHC) which assumes presence between 9am and 6pm.



- False Positive Rate (FPR) is **reduced over 30%** with CO₂ features
- Time stamp and shifted differences are the most important features

Conclusions

- CO₂ data leaks presence/absence information of specific occupants in a small multi-occupant room
- In future work, larger spaces and more diverse occupants (w.r.t gender, age etc.) can be tested