Project Reference Manual

For

Thermal Comfort Analysis Using Machine Learning

- **Manas Kala, Osaka University**
- **Betty Lala, Kyushu University**

Project Overview

Understanding and predicting thermal comfort of occupants of an indoor environment (office, class, home) is a complex problem. Thermal comfort preferences are highly personal and subjective, therefore predicting/modeling thermal comfort of a group (class, office etc.) is difficult. A naturally ventilated environment which can't be controlled and the impact of outdoor factors (temperature, rain, humidity, wind-speed) adds more complexity to the problem.

When it comes to primary school children, thermal comfort analysis and prediction becomes even more challenging. Children have higher sensitivity to temperature and lack the cognitive ability and authority to express and take corrective action (modify environment, clothing etc.). Since learning outcomes and scholastic performance are directly linked to the thermal comfort in classrooms it is extremely important to analyze and predict primary school students' satisfaction in the naturally ventilated classrooms. However, due to the complexity of this problem combined with the challenges involved in conducting surveys with primary school children as participants, only a handful of studies have attempted to do so. For example, in the global thermal comfort databases ASHRAE I and II, not a single study in India is conducted for primary school students, or even for naturally ventilated classrooms.

To the best of our knowledge, this is the first study in the Indian subcontinent to analyze and predict thermal comfort for naturally ventilated primary school classrooms using Machine Learning.

This document describes:
1. The Data for different datasets (Winter, Summer, Teacher Survey)
2. Data Distribution
3. Consolidated Research Problems
4. Model Details and Micro Problems

I. Data Description

Input Features	Output Labels
Avg. Daily Maximum Temperature Avg. Daily Minimum Temperature Indoor Temperature (During the survey) Relative Humidity (During the survey) School Classroom Time of Experiment Grade Age Gender Individual Clothing Value (Clo) Total Clothing (Total Clo) Total Clothing with Chair Day of the survey/experiments	Thermal Sensation (Vote): TSV$^{\Phi}$ Thermal Comfort (Vote): TCV$^{\Phi}$ Thermal Preference (Vote): TPV$^{\Phi}$ Thermal Environment Acceptability: TA$^{\Phi}$ Temp Satisfaction Level: TSL$^{\Phi}$ Satisfaction with Clothes: SwC* $^{\Phi}$ Modified Clothing: MC* $^{\Phi}$

Φ Some outputs can be used as input features in predicting other outputs

*****MC and SwC can be considered as both input feature and output depending upon the problem.

Relevance of Inputs

Avg. Daily Maximum Temperature, Avg. Daily Minimum Temperature, Indoor Temperature (During the survey)	Temperature directly impacts Thermal Comfort. Important to know which of these three temperatures is the most important to predict thermal comfort.
School, Classroom	Schools and Classrooms differ in geographical location, architectural layout, direction of sun, wind direction, lighting, school uniforms. Both combined and School and Classroom specific analysis is important.
Age, Grade	Sensitivity to temperature, ability to express discomfort, process the information regarding opportunities available to modify the environment, and the ability to modify clothing and environment depends on Age and Grade (Cognitive abilities).
Gender	Temperature Sensitivity, Clothing, Metabolic Activities, and ability to modify environment/clothing often depends on the Gender.
Day of Experiment	Surveys are intrusive i.e., they cause a disruption in participants life. Children tend to have shorter attention spans and lower cognition. So, it is important to see how their responses vary on different days of surveys and experiments.
Time of Experiment	As time progresses, the day gets warmer, which will impact thermal comfort perception of students.
Individual Clothing Value*	Each clothing item contributes to thermal insulation and overall thermal comfort. Ranking the importance of individual clothing items is important.

Total Clothing Value*	Sum of individual clothing values. (Plus Chair)
Modified Clothing	To test (a) Impact of Age, Gender, Grade, School, Classroom on Children's ability to take action (b) Correlation with other outputs
Satisfaction with Clothes	Combined with Modified Clothing does dissatisfaction translate into corrective action?

*Individual Clothing Values are for granular analysis and solving micro-problems. Total Clothing Value (with/without Chair) is more suited for general thermal comfort prediction. **Don't consider multiple (Consider only one of the three).**

- This is the input feature list based on the current state of data that has been encoded from paper-bases surveys. Gradually, more features will be added corresponding to classroom environment.
- Some Objective features like Avg. Temperatures are purchased from IMD while others such as Indoor Temperature and Indoor Humidity were measured using sensors/IoT devices.
- Time slots of the experiment/survey are encoded as follows:

Encoded Start Time	
<=9:30	1
>9:30 <=10:00	2
>10:00 <=10:30	3
>10:30 <=11:00	4
>11 <=11:30	5
>11:30 <=12:00	6

II. Data Distribution (Winter Data Only)

5 Schools, 14 Classrooms, 3-5 days/class, 512 unique student participants: 2038 sample size.
Surveys+Measurements + IMD weather data
The distribution of data is presented below.
Relevant for (a) Classification using Data Imbalance techniques (b) Outlier detection and removal

School	
1	543
2	298
3	544
4	514
5	139
	0
	2038

Day	
1	462
2	490
3	433
4	437
5	216
6	0
	0
	2038

Start Time	
9:30	647
10:00	647
10:30	415
11:00	185
11:30	104
12:00	40
	2038

Gender		
1	970	Male
0	1068	Female
	2038	

Age	
6	3
7	45
8	286
9	626
10	753
11	278
12	44
13	3
	0
	2038

Class	
3	704
4	738
5	596
	0
	2038

Encoded Start Time	
1	647
2	647
3	415
4	185
5	104
6	40

TSV	
-3	0
-2	98
-1	528
0	1354
1	57
2	1
3	0
	0
	2038

TPV	
-2	0
-1	9
0	868
1	899
2	262
	0
	2038

TCV	
-3	5
-2	12
-1	12
0	0
1	224
2	1209
3	576
	0
	2038

TA — Thermal Environment Acceptability

TA	
1	84
0	1954
	0
	2038

MC — Modified clothing

MC	
1	248
2	54
3	19
4	1717
	0
	2038

SWC — Satisfaction with Clothes Worn

SWC	
1	129
2	1850
3	59
	0
	2038

TSL — Class Temperature Satisfaction Level

TSL	
-3	0
-2	13
-1	37
0	0
1	341
2	1151
3	496
	0
	2038

TPV_Binary		
0	1170	Want Change
1	868	Comfortable
	0	
	2038	

MC_Binary		
0	321	Changed Clothes
1	1717	Comfy
	0	

TA_Binary		
0	84	Not Acceptable
1	1954	Acceptable
	0	

SWC_Binary		
0	188	Not Satisfied
1	1850	Satisfied
	0	

Relevance of Outputs

The scales have been modified to fit the responses for winter data and to make analysis easier.

Metric Scales

Thermal Satisfaction Vote (TSV)	
Very Hot	3
Hot	2
Warm	1
Neutral	0
Cool	-1
Cold	-2
Very Cold	-3

Thermal Environment Acceptability (TA)	
Acceptable	0
Not Acceptable	1

Modified any Clothing (MC)	
Remove a clothing Item	1
Wore an extra clothing	2
Loosen a clothing	3
No Change	4
Invalid	0

Thermal Preference Vote (TPV)	
Much Warmer	2
Bit Warmer	1
No Change	0
Bit Cooler	-1
Much Cooler	-2

Satisfaction with Clothes (SWC)	
Wearing Excess	1
Satisfied	2
Wearing Less	3

Thermal Comfort Vote (TCV)	
Very comfortable	3
Comfortable	2
Slightly Comfortable	1
Slightly Uncomfortable	-1
Uncomfortable	-2
Very uncomfortable	-3

Temp Satisfaction LeveL (TSL)	
Very satisfied	3
Satisfied	2
Slightly satisfied	1
Slightly dissatisfied	-1
Dissatisfied	-2
Very dissatisfied	-3

- Both multi-level and binary assessments of thermal comfort are considered.
- Multiple levels are collapsed into a binary scale 0: Not Comfortable and 1: Comfortable.
- Some outputs can serve as features for other outputs
- Not all levels/values in the scales are present in the dataset as some conditions are not applicable in the winter season (e.g. Hot, Very Hot for TSV)

III.Consolidated Problem Statements

The problems outlined below include aspects that have either not yet been explored or are not investigated enough. New technical approaches that are not yet (widely) used in the domain are also considered. Ideally, each of these problems will have enough contributions for a standalone paper, depending upon (a) Problem can be solved (b) Solid results with high accuracy. Multiple problems can be merged for more significant contributions or for submissions to high-impact journals.

***Feel free to explore and propose (a) New problems (b) New technical solutions**

Problem 0: Preliminary Analysis

Step 1: Exploratory Data Analysis

- Understand the data
- Identify non-linear relationships between features and labels/outputs
- Discover insights and plot interesting findings

Step 2: Dimensionality Analysis/Feature Selection

- Identify the most important features for each label
- Try using PCA and t-SNE

Step 2: Feature Relationship Analysis between Input and Output Variables

- Linear Correlation Analysis: Pearson, Distance etc.
- Do techniques like Rank and Partial Correlation improve results?
- Try non-linear correlation techniques

Keep adding the results (plots, graphs, heatmaps, etc.) to the Overleaf/Latex File with basic observations. Leave the technical writing and detailed analysis for later.

Problem 1: The curse of multiple subjective outputs

Which and how many (output) subjective metrics should be used for the assessment of thermal comfort (TC) perception of the occupants?

Need to go beyond the conventional approaches (linear correlation and PCA) used in current studies.

Datasets: Gathered Winter Data, Classroom data from ASHRAE Dataset (For Baseline Comparison)

Tasks:

- Focus on the 5-6 subjective TC metrics (TSV, TCV, TPV, TA, SwC, TSL)
- Consider Multi-class (levels of comfort) and Binary (Multiple levels collapsed into comfortable/not comfortable)
- Correlation Analysis (Outputs only): Try Several correlation techniques
    - Linear Correlation Analysis
    - Rank, Partial (Discount for correlations from other variables)
    - Non-linear correlation (if non-linear relationship exists)
- Dimensionality Analysis: PCA, t-SNE, other techniques
    - Identify and select most important Feature Variables
- Predict less important output variables using most important variable identified from the analysis (Both Binary and Multi-class classification)
    - Use SVM, RF, ANN, Deep Learning (if it makes sense)
    - k-Fold Cross Validation (standard practice)

Problem 2: Predicting Behavior of Students in Response to the Classroom Thermal Environment

1. Predict whether children are likely to modify clothing if they are uncomfortable or dissatisfied with the (a) Thermal environment (b) Clothing (c) Classroom Environment.
2. We can consider all features to predict Modified Clothing and see which feature determines the most whether children will modify their clothes.
3. Further, does Satisfaction with Clothing determine Modification → This will determine children's ability to exert authority or express themselves. Do Age, Gender, Grade play a role here? Does it vary across schools and classrooms?

Tasks

- Correlation and Dimensionality Analysis
- Binary and Multi-class Classification/Prediction using multiple techniques
- Consider micro-problems related to School, Classroom, Day, Time, Age, Grade, Gender, Clothing

Problem 3: Impact of Class Environment on Thermal Comfort Perception of Primary School Children (Light, Cleanliness, Acoustics etc.)

Assess the impact of the additional data about classroom environment on Thermal Comfort perception and occupant behavior:

Tasks

- Dimensionality analysis
- Binary and Multi-class prediction
- Consider micro-problems related to School, Classroom, Day, Time, Age, Grade, Gender, Clothing

Problem 4: Clothing and Thermal Comfort

Tasks

- Rank clothing in the order importance/weights
- Which items are least significant in school uniforms?
- How does the impact of clothing vary between boys and girls?
- Impact of chair? Impact on models if alternatives to wooden chairs are used (with cushion, etc.)
- On sports day/games day/Gym class: Students don't wear the usual Uniform. On this day they wear Jackets and trackpants. Analyze the impact of dress change on Thermal Comfort Perception.

Relevance of Clo values: Do clothing values translate directly into thermal comfort?

Clo Values are a good measure but they may not always translate well to Thermal Comfort perception. A stocking, Tie, Scarf, Cap may have very low Clo values (Or the clothing value may not be available) but it may protect the throat and chest from cold and may lead to a signficant change in thermal comfort.

- Check if this is true
- Apart from Clo based analysis, a wearing-not wearing impact analysis on Thermal Comfort perception for all clothing items can be done.

Also, what influences the choice of clothing the most?

- Min/ Max Avg. temperature of the day?
- Min/Max Avg. temperature of the previous day?
- Humidity, rain, etc?

~~Problem 5: Multi-task Learning~~

Problem 6: Improve accuracy of models using Data Imbalance techniques

Problem 7: Outlier Analysis

Problem 8: Combined Analysis with Summer and Winter Data

Problem 9: Teacher Survey Analysis

IV. Model Details and Micro-level Research Problems

i. Binary/Multi-class Classification/Prediction model should be able to accurately classify/predict the expected output/label value in comparison to the ground truth value.

- Thermal Sensation (Vote): TSV
- Thermal Comfort (Vote): TCV
- Thermal Preference (Vote): TPV
- Thermal Environment Acceptability: TA
- Temp Satisfaction Level: TSL
- Satisfaction with Clothes: SwC
- Modified Clothing: MC

For Binary Classification Collapse multiple classes of the output/label values to represent two classes:

- Comfortable/Satisfied
- Not Comfortable/Dissatisfied

ii. For each of the classification models, determine the weight/importance of the input following feature variables in determining the overall accuracy of the Classification model:

- Feature selection and Dimensionality reduction
- Dimensionality reduction using advanced techniques like t-distributed stochastic neighbor embedding.
- Comparison with PCA (Most existing studies only use PCA)

iii. Identify the impact of Schools, Time of the experiment, Day of the experiment, Age, Gender, Class/Grade on the accuracy of models and comfort levels

- Distribution of output labels (e.g., TSV/TCV/TSL) values across these parameters e.g., Age vs TSV, Gender vs TSV etc.

- Does accuracy increase/decrease because of these features?
    - Are the models more accurate for boys or girls (combined, school-wise, class-wise)?
    - Does the accuracy increase with class/grade and age (combined, school-wise)? → Impact of cognitive ability, special needs, age.
    - Does the accuracy vary across schools?
    - Difference in model accuracy between Govt and Public schools?
    - Does the accuracy change with the Time of the experiment? (Perception of Thermal Comfort is different in cold winter mornings) → Will help in determining what is the ideal time for conducting the experiments in Winters.

    - Does the accuracy change with the Day of the experiment? (Initially there might be confusion, lack of clarity, then a better understanding, finally giving way to boredom) → Will help in determining/prescribing the ideal duration for the experiments.

iv. Impact of Temperature and Humidity on model accuracies:

- Which temperature is the most important for the model (Avg. Maximum Temp, Avg. Minimum Temperature, Indoor Temperature during the Experiment)?

--