Interpolation of Sparse Indoor Temperature Data in Space and Time

Sam Peters[1,2], Joseph McLaughlin[1,2], Michal Young[1,2], Siobhan Rockcastle[1,3,4]

[1]University of Oregon, [2]Department of Computer and Information Science, [3]Department of Architecture, [4]Baker Lighting Lab

CAS

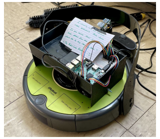
UNIVERSITY OF OREGON

Problem Statement

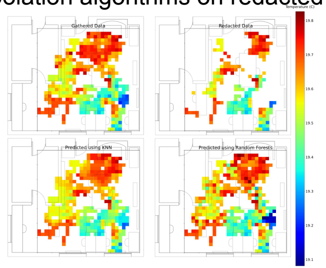
Evaluate algorithms for interpolating sparse and unevenly distributed temperature data.

Methods

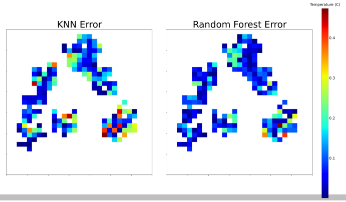
Use parasitic data collection process to collect environmental data across a number of days and times.



Redact collected data, then run interpolation algorithms on redacted data.



Compare interpolated and real values.



Results

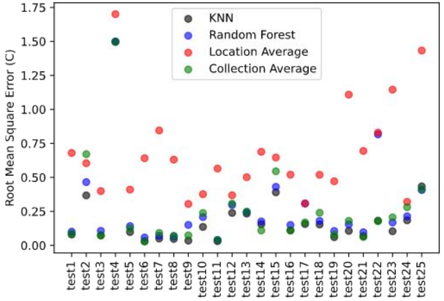


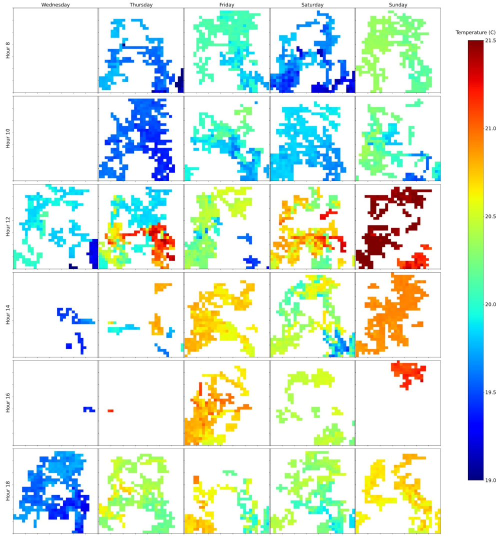
Figure: Algorithm performance across a number of tests, measured using the root mean square error metric

The k-nearest neighbors, random forest, and collection averaging methods were all able to outperform the time-based location average interpolation method. KNN, random forests, and collection average had average root-mean square errors of 0.200, 0.266, and 0.245 respectively, versus 0.688 for location average. This makes sense intuitively for this data set, given that we're working with indoor temperature data which tends to deviate less than other environment data.

During the training process, the random forest model assigned large weights to the values of nearby points, and extremely small weights to other features such as time and location, effectively degenerating to a model resembling KNN. Therefore, performance of the KNN and random forest algorithms were similar.

For this data, simple spatial inference was superior to searching for more complex spatiotemporal relationships, even when losses of data are concentrated in a few large areas.

Data



Future Directions

- Anomaly detection
- Parasitic data collection with a broader range of devices and environments
- Calibrating real world building models