Poster Abstract: Towards a Predictive Model for Improved Placement of Solar-Powered Urban Sensing Nodes

Alex Cabral
Harvard University
acabral@g.harvard.edu

Jim Waldo
Harvard University

Amy V. Mueller
Northeastern University

***Abstract*—In a world driven by data, cities are increasingly interested in deploying networks of smart city devices for urban and environmental monitoring. To be successful, these networks must be reliable, low-cost, and easy to install and maintain—criteria that are all significantly affected by the design choices around power and can seemingly be satisfied with the use of solar energy. However, solar power is not ubiquitous throughout cities, making it difficult to know where to place nodes to avoid charging issues and thus potentially increasing maintenance costs. This abstract describes the development of a machine learning model that predicts whether any arbitrary location in a city will have solar charging issues. Using data from a large-scale real-world solar-powered sensor deployment in Chicago, Illinois and open data about building location and height, the binary classification model outputs the probability of adequate solar charging at a node location with 77% accuracy on the held-out test set. This work lays the foundation for those deploying future solar-powered urban sensor networks to have more confidence in the reliability of their chosen node locations.**

I. Introduction

As the global urban population continues to grow, cities are increasingly interested in monitoring urban processes such as vehicular traffic, and environmental harms including air pollution and noise, to help cities grow in a healthy and sustainable fashion [1], [2]. The lowering cost of sensing infrastructure has encouraged city officials, researchers, and urban residents to use large-scale, low-cost sensor networks to collect data, monitor hyperlocal phenomena, and inform policy to help transition to becoming smart cities [2], [3].

Prior research indicates that to be successful an urban environmental sensor network must be reliable, maintainable, and low-cost [3], [4]. One key feature of an urban sensor network design that helps achieve these goals is a reliable power source, which provides for real-time node monitoring and ongoing operation. Nodes must be continuously running to collect data over time, yet many outdoor urban spaces are not equipped with accessible wired mains [5]. Solar power is the most ubiquitous form of renewable energy for sensor networks, and will remain prevalent in the coming years because of its relative low-cost and ease of scalability and maintainability.

Despite the existence of highly accurate models for solar insolation as a function of latitude and longitude, urban form can make it complicated to infer the effective solar insolation received at any given location. Former research has examined the effect of buildings on the amount of light in urban areas based on the position of the sun in relation to buildings and the shadows they cast [6]. However, this tool only has computed shadow data for five US cities – NYC, Chicago, Los Angeles, Boston, and Washington, DC – and is primarily useful for flat terrains so will not generalize to hilly cities [6].

Specifically around urban sensor networks, Dehwah et al. [7] evaluated the performance of a traffic monitoring sensor network in a desert city. The authors found that dust storms and building shadows had a drastic effect on solar charging, each sometimes resulting in more than 90% loss in solar panel output. However, the authors do not do a deep analysis into the locations that were most affected by shadows to determine how the issue may be identified or prevented in future deployments.

This poster abstract presents a machine learning model that uses data from 106 solar-powered low-cost sensor nodes deployed in Chicago, USA between July 1, 2021 and June 30, 2022 [3]. Based on the network loss of data due to limited solar charging between October and March and open data about the buildings near the sensor node locations, a Random Forest classification model is developed and trained to predict if a single node location will experience solar charging issues. The model has an accuracy of 0.77 and precision $(TP/(TP+FP))$ of 0.85 on the held-out test set, indicating that the model can identify node locations that will charge successfully throughout the year with high probability.

II. Methods

The large-scale solar-powered sensor network, described in further detail in [3], [8], was designed and deployed to collect air quality data across Chicago. The network was comprised of 118 sensor nodes at 110 unique locations including 106 bus shelters and 4 regulatory monitoring stations. Each sensing node was outfitted with a rechargeable 2000 mAh lithium polymer battery and a $10{\times}13$ cm Voltaic Systems P126 6W solar panel. On average, the device drew 4mA current over a 24 hour period, allowing the battery to power the sensing node, including communications, for approximately 15 days at a sampling rate of 5 minutes.

In October 2021, one of the devices stopped charging because the sun was no longer reaching the solar panel due to the change in the solar position and the node being surrounded

979-8-3503-7025-6/24/$31.00 ©2024 IEEE
DOI 10.1109/IoTDI61053.2024.00023

by skyscrapers. A power-saving mode was implemented to ensure the network still collected useful data throughout the winter months. Power-saving mode (PSM) was initiated when the battery power level fell to 15% or less of its total capacity then turned off when the battery power level had recharged to at least 40%. PSM information was used to identify nodes that had charging issues in the winter months, with nodes that had power throughout the entire winter set to a class of 1 and nodes that experienced PSM set to a class of 0. OSM (Open Street Maps) Buildings [9] was used to gather data about buildings surrounding the nodes. Specifically, we examined the distance and height of the closest building and the number, mean height, median height, and maximum height of buildings within 25, 50, 100, 250, and 500 meters of each sensor node, which were all used as predictors for the model.

The sensor network and OSM Building data were combined into a dataset of the 106 bus shelter locations [1], 44 of which experienced PSM and thus were in class 0. The data were log normalized and scaled then split into training, validation, and test sets at a 74%, 16%, 20% stratified split to preserve the class imbalance. Several models were trained and tuned on the training and validation sets and different combinations of predictors were used to prune features. The best performing model was then run on the held out test set.

III. Results and Discussion

The best performing model is a Random Forest model trained and tuned on input features that include data about the closest building and those greater than 100 meters away. The results of that model on the held-out test set are shown in Table I. The feature importances show that the closest building data explain nearly 30% of the variance, which is logical given the likelihood of extremely close buildings blocking the path to solar charging.

TABLE I
ACCURACY RESULTS FROM THE RANDOM FOREST MODEL TO PREDICT POWER CAPABILITIES ON THE HELD OUT TEST SET USING INPUT FEATURES FOR THE CLOSEST BUILDING AND THOSE ≥ 100 METERS AWAY.

Test Acc.	True Pos	False Pos	True Neg	False Neg
0.77	0.79	0.25	0.75	0.21

The model results show promise for improved solar-powered node placement in urban areas to help limit the costs of maintenance and ensure reliability of deployed urban networks. However, a number of limitations arise in understanding the utility of these proposed quality metrics in other urban settings. One of the main limitations is the reliance on open crowdsourced data to determine the location and height of buildings in the city. As with many open crowdsourced datasets, these data are not completely accurate or up-to-date [10]. Open data also cannot capture every potential issue, such as a trees that may be blocking nodes from the sun.

Another limitation is that the data come from just one deployment in one city, raising the question of how well Chicago generalizes to other cities. Chicago is at a latitude of 41.88 degrees, where the sun is at a maximum zenith angle of 24.5 degrees above the horizon on the winter solstice [11]. The five most populous latitudes are between the 22nd and 27th parallel north [12], which are closer to the equator and have an maximum solar zenith angle of 42 degrees on the winter solstice [11]. Nevertheless, a number of highly populated cities exist at or above the 42nd parallel north, including London, Moscow, Harbin, Toronto, and much of Western Europe. New York City and Beijing are also located at nearly the same latitude. Future work to understand the generalizability of similar models will be critical to leveraging these insights.

[1]Calibration nodes were on rooftops so excluded from the analysis.

IV. Conclusions and Future Work

This abstract introduces a predictive model for evaluation of solar resource availability at a given location, which can be used to improve the placement of solar-powered sensors in cities. The model uses open data about urban form and provides promising results in classifying singular locations for adequate or poor charging capabilities. However, the lack of accurate open data points to a need for new methods to obtain up-to-date urban data. Furthermore, there is a huge need for further real-world deployment and data collection to improve the understanding of solar charging issues for urban environmental sensor networks. With data from additional cities, the model can expand to use solar zenith information and become more generalizable, offering use for those deploying urban sensor networks in more cities around the globe.

References

[1] United Nations, "68% of the world population projected to live in urban areas by 2050, says UN," 2018.

[2] C. C. Okafor *et al.*, "A bibliometric evaluation and critical review of the smart city concept – making a case for social equity," *Journal of Science and Technology Policy Management*, Jan 2022.

[3] M. I. Daepp, A. Cabral, *et al.*, "Eclipse: an end-to-end platform for low-cost, hyperlocal environmental sensing in cities," in *2022 21st ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, pp. 28–40, IEEE, 2022.

[4] R. N. Murty, G. Mainland, *et al.*, "CitySense: An Urban-Scale Wireless Sensor Network and Testbed," in *2008 IEEE Conference on Technologies for Homeland Security*, pp. 583–588, May 2008.

[5] J. Adkins, B. Ghena, *et al.*, "The signpost platform for city-scale sensing," in *2018 17th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, pp. 188–199, IEEE, 2018.

[6] F. Miranda *et al.*, "Shadow accrual maps: Efficient accumulation of city-scale shadows over time," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, pp. 1559–1574, March 2019.

[7] A. H. Dehwah, M. Mousa, and C. G. Claudel, "Lessons learned on solar powered wireless sensor network deployments in urban, desert environments," *Ad Hoc Networks*, vol. 28, pp. 52–67, 2015.

[8] M. I. Daepp, A. Cabral, *et al.*, "The "three-legged stool": Designing for equitable city, community, and research partnerships in urban environmental sensing," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–19, 2023.

[9] OSMBuildings, "Osmbuildings/osmbuildings: 3d building geometry viewer based on openstreetmap data."

[10] F. Syafiq, H. Aris, H. Ismail, and S. Yusof, "Automated update of crowdsourced data in participatory sensing: an application to crowdsourced price information," in *International conference on collaborative human intelligence and crowdsourcing applications*, pp. 1–7, 2016.

[11] T. Skilling, "Dear tom,in illinois, how high in the...," Nov 2000.

[12] A. Rae, "Mapped: The world's population density by latitude."