WEES-2020 Special Issue

J. Indian Chem. Soc.,
Vol. 97, No. 10b, October 2020, pp. 1861-1866

Optimal placement strategy for indoor environment monitoring using portable cost-effective devices

Praveen Kumar Sharma*, Tamal Mondal, Ananya Mondal, Anjali Ray, Sandip Mondal, Subrata Nandi, Tanmay De and Sujoy Saha

Department of Chemical Engineering, National Institute of Technology Durgapur, Durgapur-713 209, West Bengal, India

E-mail: praveencs54@gmail.com

Manuscript received online 25 May 2020, revised and accepted 24 September 2020

Monitoring the pollutant concentration is the initial and important measure to control the indoor air quality. We developed an Environment Monitoring Device (EMD) consisting of pollutants, particles, and meteorological sensors with controllers. However, monitoring each room of a building is infeasible as it increases the deployment cost of EMDs linearly. Such an issue can be resolved by deploying the devices sparsely in some rooms and estimating the air quality of other rooms using machine learning techniques. In this work, an optimal placement strategy of air quality monitoring devices has been proposed by inferring the optimal locations of devices such that the estimation accuracy is greater. Here, two scores have been used: Temporal and Spatial, to capture the impact of time and space variations. Finally, we use different clustering algorithms to get the optimal locations for placing the EMDs. A comparison of estimation using the different number of devices has also been depicted. The results have been presented in several ways and validated the proposed methodology.

Keywords: Indoor air quality, sensor placement, environment monitoring device, temporal score, spatial score, clustering algorithm, optimal placement.

1. Introduction

Degradation of air quality has dragged the attention of researchers from different disciplines. From the survey[1], it is evident that the indoor air quality is more polluted than that of outdoors. Some of the indoor air pollutants may affect the health of the occupants both physically and cognitively. As an instance, deadly diseases like pneumonia, stroke, ischaemic heart disease, and many more are caused due to indoor air pollution. As a person can spend 90% of the time indoors[2], therefore, real-time monitoring of indoor air quality should be the primary concern in order to maintain a healthy environment. A survey by Martins and da Graca[3] discussed the different aspects of $PM_{2.5}$ indoors and outdoors with their sources and impacts. In our study Sharma *et al.*[4], a sensor-based Environment Monitoring Device (EMD) has been developed to measure the indoor air quality specifically $PM_{2.5}$ and CO_2. It has been observed that in an indoor environment i.e. classroom with a certain number of occupant, the pollution level crosses the prescribed level of acceptance after a certain period of time. Now, the question lies in designing appropriate mechanisms that would optimally locate *n* number of EMDs for any given number of indoor locations. Large numbers of EMDs are required to monitor the entire building as each room has a different ambient as per the activity going on in that room. Placing the EMD in each room is not feasible as it would increase the cost linearly with extra overhead of maintenance. Besides, it also requires human intervention to maintain the infrastructure for the devices. Moreover, if multiple buildings in a particular range is to be monitored then, the problem of placing the limited number of EMDs would be the matter of concern for efficient monitoring of pollution level indoors. On the other hand, for an infinite number of available EMDs, one should also carefully analyze the optimal number of EMDs to be placed in desired locations in order to minimize the deployment cost. Considering such challenges, we design an optimization model by exploiting the temporal and spatial aspects for not only evaluating possible locations for limited EMDs indoors but also for evaluation of the placement of an optimal number of EMDs into desired locations if they are plentiful. Our contribution lies in two phases: *(i) For any limited number of EMDs, we estimate the optimal locations where the EMDs are to be placed such that the accuracy of air quality estimation is high in unmonitored locations. (ii) If the number of available EMDs is abundant then we estimate the optimal number of EMDs to be placed to monitor the air quality of the desired locations*.

The rest of the paper is organized as follows: In Section 2, several state of the art sensor placement techniques have been studied, proposed in the past literature. The proposed work has been explained in Section 3, whereas Section 4 describes the experimentation part and Section 5 explains the results and discussion. Finally, Section 6 concludes our work and provides some future research directions.

2. Related work

Some existing work explains the placement of sensors in different aspects. Castello *et al.*[5], Rackes *et al.*[6] use Geostatistical Analysis and Monte Carlo theory followed by variogram and kriging. The mean variance of the kriging is calculated and stored as the optimization variable and the minimum mean variance point provides the optimal location for the sensor node. Shi *et al.*[7] proposed a method for optimal sensor location using the eigenvector sensitivity method. Waeytens and Sadr[8], Papadopoulou *et al.*[9] proposed a strategy for optimal placement of indoor air quality sensors using computational fluid dynamics (CFD). Yoganathan *et al.*[10] used a clustering algorithm, information loss approach, and Pareto principle to derive optimal placement strategy. None of the above mentioned works have used both the temporal and spatial features in an indoor air quality perspective. Unlike the above-mentioned approaches, the current study proposes a data-driven unsupervised approach for the placement of an optimal number of EMDs indoors. The proposed technique includes temporal features like meteorological parameters and spatial features which include the area, volume, air exchange rate, etc. of any room, for evaluation of Temporal and Spatial scores. The evaluated scores are then classified using clustering techniques to obtain the optimal placement of EMDs.

3. Methodology

In this work, we have proposed a method to find the optimal rooms where the devices are to be placed to estimate the pollutants' concentration of an unmonitored room, given the number of available EMDs. Here, our objective is to optimally place the EMD, i.e. to install the EMDs in some of the rooms such that complete building can be covered. The problem can be stated as follows: Given, *n* different rooms with some spatial and temporal features. If the number of available EMD is *n*, then how these EMDs are to be placed in optimal rooms such that the real-time estimation of the pollutants can be done with greater accuracy. Moreover, what is the optimal number of EMDs required to monitor the rooms?

We gather the outside meteorological and AQI data for Temporal Scoring and classroom specifications data for Spatial Scoring. We have divided the data sets into two forms, (a) Temporal: The parameters which are time-dependent and vary with time and (b) Spatial: The parameters related to space and do not vary with time.

3.1. Temporal Scoring of rooms using meteorological data :

For temporal scoring, the correlation between all the features has been analyzed. Indoor, the pollutant CO_2 and particle $PM_{2.5}$ are relatively predominant. Hence, the temporal score is the measure of the significance of these two pollutants in the form of available features, which are Wind Speed (WS), Wind Direction (WD), Temperature (T), and Relative Humidity (RH) (Table 1).

Table 1. List of temporal and spatial features used for calculating the temporal and spatial scores

Temporal features:	Wind Speed, Wind Direction, CO_2, $PM_{2.5}$, Temperature, Relative Humidity
Spatial features:	Volume of the Room, Distance Vector, Floor, AER, Age Group

3.1.1. Feature description:

Meteorological data consists of Temperature, Humidity, Wind Speed, Wind Direction, etc. have great importance in predicting air quality in a naturally ventilated environment. The indoor environment is highly susceptible to outdoor weather conditions, and it has a direct impact of wind direction as well as wind direction in a combination of outdoor air quality. Moreover, the indoor parameters such as infiltration, exfiltration, and airtightness of the room explicitly depends upon these meteorological parameters. In order to estimate the values of the pollutant, CO_2 and particles $PM_{2.5}$ inside each room, regression models are applied. The linear regression model is used for its simplicity, but the result was not significant because it cannot fit the data sets in a straight line. Therefore, polynomial regression is used to avoid such under-fitting situations. Suppose, C_i^j, P_i^j are the *i*-th coefficient of correlation with CO_2 and $PM_{2.5}$ respectively for *j*-th room and the total number of components is *k*. So, the temporal score is calculated as in eq. (1).

$$TS_j = \frac{\sum_{i=1}^{k} C_i^j P_i^j}{\sum_{i=1}^{k} C_i^j \sum_{i=1}^{k} P_i^j} \tag{1}$$

3.2. Spatial scoring of rooms using spatial data:

The spatial scoring of rooms relies upon the features i.e. Volume of the Room, Distance Vector, Floor, AER and Age Group (Table 1). The description of each of the features are given in Section 3.2.1.

3.2.1. Feature description:

The basic features of a room have been selected to get the spatial score which includes *Volume of the room:* for detection of the pollutants' flow inside the room, *Floor:* to get the variations to height, and *Age group:* as the exhalation rate of occupants varies with age of a person (especially students). *Distance vector* is the measure of the diversity of the ambiance using the euclidean distance. If two rooms is of similar properties but the distance between the rooms are greater, this vector helps to get the correct cluster. *The Air Exchange Rate (AER)* is the rate of air inflow and outflow, which affects the air circulation of the room. It can be obtained using CO_2 concentration of the room with equilibrium analysis. A significant relation has been discussed between the window opening, the indoor temperature, and the resulting airflow rates, by La-zović *et al.*[11]. The equilibrium analysis has also been used by Paneras *et al.*[12] which has been shown in eq. (2) which shows the relation between the obtained indoor CO_2 concentration and the airflow. [a]http://energy-models.com/ventilation-infiltration-exfiltration

$$[C_{CO_2}] = B_0 V^{b_1} \tag{2}$$

where, $[C_{CO_2}]$ is the concentration of CO_2 and, *V* is the airflow rate. B_0 and b_1 are the coefficients. These can be calculated using obtained concentration of indoor and outdoor CO_2 concentrations. For spatial scoring, the features of the rooms are taken into consideration such as volume of the room, floor of the room, direction of the air flow and Air Exchange Rate (AER). AER is used to measure the rate at which the outdoor air replaces the indoor air within the room as depicted in eq. (3).

$$\text{AER} = \frac{1}{t} \ln \frac{(C(0) - C_0)}{(C(t) - C_0)} \tag{3}$$

where, $C(0)$: initial CO_2 concentration in the classroom. $C(t)$: CO_2 concentration at time *t* and, *t* is the time interval between two consecutive data points.

An equation is formed using all the spatial features as mentioned above in Table 1 for the calculation of the Spatial Score. Suppose α_i^j is the values of the *i*-th spatial parameter of *j*-th room. The Spatial Scoring of the *j*-th room can be given as depicted in eq. (4).

$$SS_j = \frac{1}{k+1} \Pi_{i=1}^{k} \alpha_i + d_c \tag{4}$$

where *k* is the number of available spatial features except distance vector. Here, α_i are the value of the *i*-th spatial feature and d_i is the distance vector from the origin and can be calculated using the eq. (5).

$$d_c = f(c, n, d_{c-1}) \tag{5}$$

where *c*: current room, *n*: neighbor.

Here, $d_0 = d_s = 0$ {s: source}

On calculating the TS and SS for each room, these points are projected on a 2-d space. We apply clustering algorithm to obtain the centroid of the clusters, which are then designated as the rooms where the EMD is to be placed.

3.3. Clustering:

After getting the values of the TS and SS as mentioned in the previous section, we have used the clustering algorithm. Here, clustering is to group the rooms in *k* number of clusters according to the temporal and spatial features. Here, *k* is the number of clusters to be formed, which again equals to the number of available EMDs. In our study, clustering techniques like *K*-Means, *K*-Medians, *K*-Medoids, and *X*-Means have been used.

4. Experimentation

The experiments have been done by collecting the data from different buildings of our institution. The reliability of the collected data has been analyzed followed by the data preprocessing and feature extraction. The results have also been analyzed and all the experimentation process has been discussed in the further subsections.

4.1. Data acquisition:

The data from 30 different rooms of our institute has been collected from different buildings, as shown in Fig. 1. EMDs are placed in all the rooms to collect the air quality data. These rooms are of different buildings with different volume, ventilation rates, etc. It is not feasible for us to place the EMDs in each room as it is not cost-effective. We have placed the EMDs in different rooms so that one box is placed in front of the blackboard, and the second box is placed at the back of the classroom, and the third one is placed outside the classroom. Meteorological data are crawled from a commercial website. Apart from this, we also collected the air quality data from Air Quality Monitoring Stations at different locations throughout Durgapur city.

4.2. Reliability of the collected data:

To ensure the reliability of the collected data, we have used different calibration techniques. In our work, Sharma *et al.*[13], an idea of hard and soft calibration techniques, has been depicted to calibrate the developed EMDs. We have used zero air calibration techniques to set the zero reading of the sensors by purging N_2 gas inside a vacuum chamber and placing the EMD inside the chamber. The external calibration is also carried out using an Air Quality Monitoring Station (AQMS) placed by the government. The calibrated EMD is then used for further calibration of the EMDs.

5. Results and discussion

The data samples collected from the rooms, as depicted in Table 1 then used for training and getting the coefficients. Then these coefficients are used to compute the TS and SS. After calculating the scores, these scores are plotted in a 2-d space. The number of EMDs are then fed as input to the clustering algorithm, and the classrooms are clustered accordingly. The obtained cluster heads or centroids are then selected as the optimal classroom to place the EMD. The results obtained by the TS and SS are plotted on a 2-d space. The results have been presented in different aspects, such as selecting the rooms of different buildings intentionally. It results in the validation of the proposed algorithm as we select the rooms from different buildings, so the cluster heads should be the rooms of different buildings. The results have been presented in different ways as follows:

1. Rooms of different building: We have selected 4 different buildings from our institution and selected different rooms from each building. Suppose, the room is denoted by R_i^j i.e. *i*-th room from *j*-th building and $j \in 1, 2, ..., 15$ and *j*-th 1, 2, 3, 4. Then after calculating the Temporal as well as Spatial scores we have clustered the rooms and we have taken the no. of EMDs same as the no. of buildings which is 4. The EMDs and buildings are kept the same to validate the pro-

Fig. 1. Front view of the selected buildings of our institution with their locations on Map.

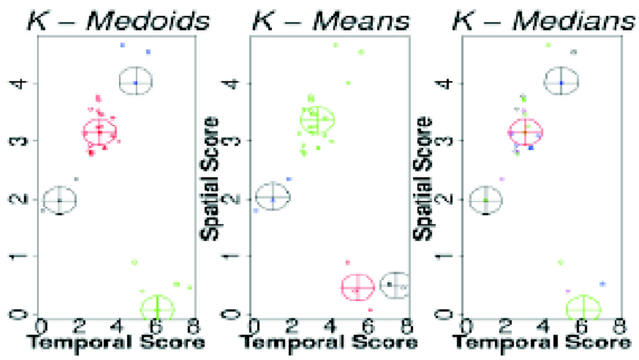

Fig. 2. The clusters of different rooms of different buildings using all the three algorithms.

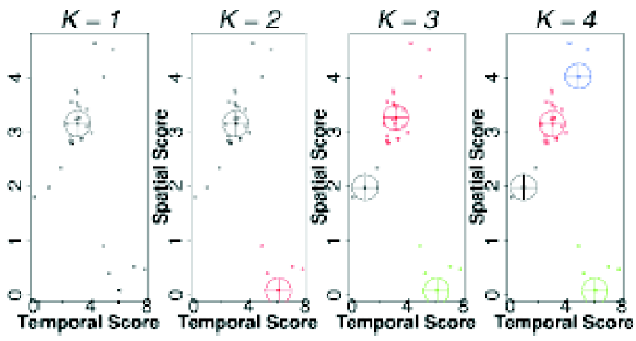

Fig. 3. The clusters formed when different rooms of the same buildings are taken with different number of EMDs available i.e. K=1, 2, 3 and 4.

posed algorithm and after clustering the TS and SS with K=4, we obtained 4 rooms as centroids as depicted in Fig. 2. Here the *k*-medoids provide the acceptable rooms as it provides the centroids as the points among the existing. So, we presented the results obtained from *k*-medoids algorithm for clustering. It has been observed that the obtained centroid rooms R_i^j have the j as the super-set i.e. the cluster head is having different j each. Hence, it can be said that our proposed algorithm works well in the first set-up but, *k*-medoids provide the centroids as a point which are among the existing.

2. Rooms of the same building: In this set-up, we have selected different rooms of the same building. Here we have selected 15 different rooms of the same building (b2). The TS and SS of these rooms have been calculated, and clustering is done. The number of EMDs varies between 2 to 4.e obtained result has been shown in Fig. 3, which describes the cluster heads for k=1, 2, 3, and 4. Here the estimation result of the final number of EMDs has also been depicted, which results in higher accuracy on increasing the number of EMDs. Here, the results describe the clusters formed using a different number of EMDs, but for K=4, a single cluster formed with only two rooms that are not feasible, so the maximum number of EMDs should be three. In this building only three floors are there so three EMDs can be placed on three floors as depicted by the cluster heads which also verifies our proposed methodology.

3. Same room with multiple trials: This test case mainly deals with the validation of our proposed method as if the same room is taken for the experiment, the optimal number of EMD should be one. We have applied our proposed method to get the optimal room with 1, 2, 3, and 4 EMDs. Here we obtain a good estimation accuracy using a single EMD while in above-mentioned cases, the estimation accuracy with a single EMD is very low.

4. Comparing the intuitive and proposed room selection: There may be a case where the intuitive selection of EMDs with optimal rooms can be made. To prove the acceptability of our work we have placed the EMD in an intuitively selected room and estimate the AQI of other rooms. Then, we have placed the same number of EMDs in the optimal rooms selected by our proposed method. We have compared the results of the estimation of air quality, and we observe that the proposed method provides good accuracy, as depicted in Fig. 4. We have intuitively selected a room in the center point of the building, and the proposed method suggests a different room. Here, the estimation is done using a machine learning technique, and when the system uses the intuitively

Table 2. Comparison of estimation accuracy by deploying the EMDs in the rooms obtained by the proposed algorithm and the intuitively selected rooms

	Different building (All buildings)				Same building (Building-1)			
	K=1	K=2	K=3	K=4	K=1	K=2	K=3	K=4
Intuitive accuracy (%)	70	76	81	89	72	85	90	92
Proposed accuracy (%)	77	79	83	92	78	87	92	93

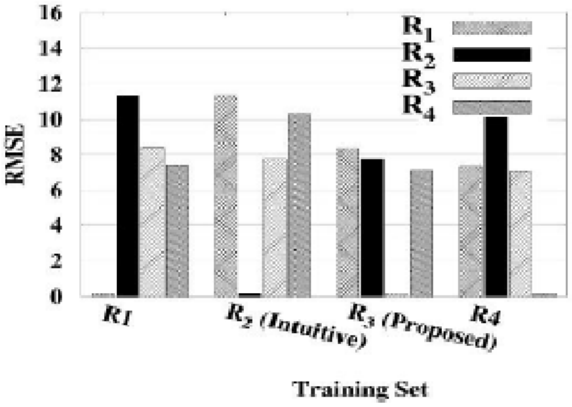

Fig. 4. RMSE plot of the estimation of one monitored and three unmonitored rooms including intuitive and proposed method.

selected room, the Root Mean Square Error (RMSE) is high compared to the one suggested by the proposed method. Moreover, Table 2 depicts the accuracy score with different values of *k* (EMD).

5. Optimal number of EMD: In this case, the number of EMDs which should be used to get the acceptable estimation accuracy is estimated. The intuitively selected room and room selected by the proposed method have estimated the air quality of unmonitored rooms by placing and varying the number of EMDs. By selecting 1, 2, 3, and 4 EMDs individually, we have calculated the accuracy using our proposed method using clustering techniques such as *k*-means, *k*-medians, and *k*-medoids. We then used *x*-means clustering algorithm to get the optimal value of *k* which is the number of EMDs in our scenario. We have to initialize the minimum number of centroids by the minimum number of EMDs (*Emin*). Then *x*-means is applied and we obtain the optimal possible clusters using *Emin*. If the clustering is not possible then the number of clusters will be *Emin* + *x*, with the least possible *x*.

6. Conclusion

In this work, we have proposed a method to place the EMDs in indoors optimally. We have used temporal as well as spatial features of rooms and calculated Temporal and Spatial scores considering the temporal and spatial features, respectively. These scores are then used in 2-d space, and some clustering techniques are used, such as *K*-means, *K*-medians, and *K*-Medoids, to cluster the rooms based on the temporal and spatial scores. Finally, the rooms which are the centroids of the clusters are marked as the optimal room. The estimation of the AQI of rooms is carried out, and we have achieved a good accuracy of 93% with optimal numbers of EMDs.

Acknowledgements

The authors would like to thank the (a) Council of Scientific & Industrial Research (CSIR), India, a premier national R&D organisation, (b) Project IntAirSense funded by Department of Science & Technology (DST), West Bengal, India, and (c) Project CityProbe funded by IMPacting Research INnovation and Technology (IMPRINT) India for funding this research work in parts.

References

1. K. Hess-Kosa, "Indoor air quality: the latest sampling and analytical methods", CRC press, 2018.
2. Indoor air quality. URL: https://www.eea.europa.eu/signals/signals-2013/articles/indoor-air-quality, 2017.
3. N. R. Martins, *et al.*, *Sustainable Cities and Society*, 2018, **42**, 259.
4. P. K. Sharma, *et al.*, "On detecting acceptable air contamination in classrooms using low cost sensors", in: 9th International Conference on Communication Systems and Networks (COMSNETS), 2017, pp. 484. IEEE.
5. C. C. Castello, *et al.*, "Optimal sensor placement strategy for environmental monitoring using wireless sensor networks", in: 42nd Southeastern Symposium on System Theory (SSST), 2010, March, pp. 275. IEEE.
6. A. Rackes, *et al.*, *Science and Technology for the Built Environment*, 2018, **24(2)**, 188.
7. Z. Y. Shi, *et al.*, *Journal of Engineering Mechanics*, 2000, **126(11)**, 1173.
8. J. Waeytens and S. Sadr. *Building and Environment*, 2018, **144**, 184.
9. M. Papadopoulou, *et al.*, *Journal of Computing in Civil Engineering*, 2016, **30(2)**, 04015024.
10. D. Yoganathan, *et al.*, *Energy and Buildings*, 2018*,* **158**, 1206.
11. I. Lazovi, *et al.*, *Thermal Science*, 2016, **20**, S297.
12. G. Panaras, *et al.*, *Sustainable cities and society*, 2018, **39**, 126.
13. P. K. Sharma, T. De and S. Saha, "IoT based indoor environment data modelling and prediction", in: 10th International Conference on Communication Systems & Networks (COMSNETS), 2018, January, pp. 537. IEEE.