Anadolu Psikiyatri Dergisi || ISSN: 1302-6631 || Volume 25; Number 4
https://doi.org/10.5281/zenodo.13341530

Big Data Optimization for Pollution Free Path Selection Using Wireless Sensor Network

Dr.K.Preethi [#1], Dr.L.Bapitha[*2], Dr.A.Backia Abinaya *[3],

1. Assistant Professor, Department of Computer Science and Engineering, University College of Engineering, Pattukkottai,614701, India.

2.Assitant Professor, Department of English, Anna University regional Campus, Coimbatore.641046, India.

3.Assistant Professor, Department of Electronics and Communication Engineering, St Joshep College of Engineering, Thanjavur.613006, India.

Corresponding Author: K.Preethi

ABSTRACT

Big data is most widely used platform for many environmental problems. Every second sensors predict data and need efficient infrastructure to process the data. It's one such technology, big data can store and process huge data efficiently. In this paper wireless sensor networks are used to monitor the air pollution in that area. The sensor data is stored in centralized server using big data hadoop system and process for optimal path selection. Air pollution is major health hazard which people cannot stop breathing a second Inspite of bad air. While travelling some people may be affected with cardiac problems, breathing troubles by inhaling polluted air. It is necessary to choose pollution free and also optimal path for travelers. Sensors sense the carbon monoxide (CO) level, sulphur dioxide (SO_2) and temperature. The sensed data are processed in two phases. In the first phase, the sensed data is preprocessed using SHA-512 hash algorithm to remove the data redundancy. Big data HDFS map reduce techniques is used here. Further features are extracted using support, confidence and frequently occur pattern recognition. The entropy is used to manage the support and confidence. Here, K-means algorithm with normalization is used to cluster the related data using entropy values. Second phase consist of checking pollution level and choosing the optimal path in the Google map. Improved-Adaptive Neuro-Fuzzy Inference System (I-ANFI) Algorithm is used to identify the pollution level of the area and optimal path is selected using ant colony optimization with genetic (ACOG) algorithm. The performance of various algorithms is compared in result section. Our proposed algorithm proved its better performance compared to existing algorithms.

Keywords: air pollution, SHA-512 hash algorithm, K-means, normalization, fuzzy, ant colony optimization, genetic algorithm.

1. Introduction

The human life is possible on earth only with clean air inhalation. Pollutants in air are invisible and it affects the health of people severely. This degrades the development of society affects the sustainable growth. Day by day more people losses their life due to air pollution-based diseases. The pollution also affects the normal life by damaging agriculture, acid rains, green house affects, ozone depletion. These all-harmful effects that badly affect quality air in the environment. There are more death in Britain, India due to smoke. Forest is fully destructed in West Germany due to acid rains. In Delhi during 2018, 2019, it was affected severely by fog and smog. Most of the people severely affected with respiratory problems in winter. Due to importance of human survival, many countries come forward in research on air quality monitoring and mitigation measures. There are few factors to determine path selection based on air quality imp acts such as wind direction, speed, motion in atmosphere, distribution of temperature, pollutants in air etc. [1]. Great challenge is to overcome chemical and physical reactions of these factors.

Complex networks are available in computer science engineering and human society. This network helps to identify the hidden rules from large number of the complex network system. Some applications of complex systems are protein network system [2], metrological communication networks [3], interpersonal systems [4], and gene regulatory system [5]. Multilevel air quality monitoring system is introduced in many countries.

Pollutants are major cause for air pollution diseases. Some pollutants like nitrogen dioxide (NO_2), carbon monoxide, carbon particles, ozone, Sulphur dioxide (SO_2) are considered as primary pollutant [6]. Particulate matters present in the air are another type of pollutants. Particles measured less than 2.5μm and 10 μm diameters cause various health issues and diseases [7]. Today, many metro cities are ready with their own air quality measuring stations to see their city pollution level every day and hour. Though there is a concern over measuring the pollution level around the every human population and give information regarding safe travelling route [8], [9]. The conventional method to monitor pollution is located in fixed place in large geographical area.

The sensor-based monitoring system has several challenges to address. Investment must me large for deployment, nearby environmental factors are considered. The roads, industries, few distances play a major role in measuring pollution levels. Therefore, it is huge demand for cheap monitoring of air quality prediction system. From 2016, there are more gas sensor manufactures in the market. SGPC10, BME680 are some leading sensors with small size and low consumption of power. They measure humidity, pressure, temperature etc in the area [10].

In our proposed work we have used multiple static sensors [11] to effectively monitor the air quality. This sensor provides continuous information to the network. Then the collected data is stored in cloud server. Then the hash algorithm uses to label and split the sensor data. K-means clustering is performed with normalization to reduce data redundancy in the collected data. The main contribution in this work is as follows:

1. We place static sensors with mobile IoT sensors to predict air pollution data.
2. Neuro fuzzy based system is used to analyze the level of pollution
3. Hybrid ant colony with genetic algorithm is used to select the optimal path.

The rest of our article is organized as follows: Section 2 presents the related research work on different air quality measurement and optimization methods. Section 3 describes data preprocessing phase. Section 4 explains proposed models and algorithms. The experimental setup and results are presented in Section 5, and concluded the work with section 6.

2. Related Research Work

In this section detailed description various contemporary algorithms and methods used are described. Recent research based on our proposed system is discussed.

A. Features Based Approaches:

The particulate matter (PM2.5) is predicted using Markova method using data with time series [12]. Most widely used hidden Markova model has memory limitation while processing the data. It has wide prediction drawbacks in tracking temporary dependent data. This problem was addressed by using hidden semi-Markova model (HSMM) for predicting PM2.5 level in the geographical area. Feature reduction technique using classifier is discussed in [13]. Support vector machine based hyper plane classifier is used to predict PM level in air. Further cuckoo search optimization is used for improving classification quality [13]. The different neural network-based approach is explained in [14]. Wavelet neural, fuzzy neural, least square based support vector approach, machine learning based network models are used in predicting the PM2.5 parts present in air. The wavelet neural models produce good accuracy and outcome when comparing with other techniques. Hourly air quality prediction is in multiple patterns are determined in [15]. The ensemble classification and regression methods are used to train the cluster models of data. Adaptive learning models in [12],[13],[14],[15] capture the long-term dependent data from all sources. It also introduces stand-alone model to analyze pollutants PM2.5,10,NO_2 for 24 hrs.

B. Deep Learning Based Models:

The concentration of particulate matters are forecasted using deep learning models by recent researchers. China urban data uses novel algorithm to predict missing data and pollution prediction done using deep learning models [16]. Industrial air pollution is predicted using deep neural network [17] . Future air quality values are predicted using Long Short-Term Memory (LSTM) by using IoT devices in smart cities [18]. Deep framework for air learning is adopted by authors in [19] for using the unlabeled information of spatio-temporal data for increasing the performance of prediction.

C. Wireless Network Models:

Mobile based wireless sensor networks are used to give pollution free maps based on measuring values of sensors and other pollution values [20]. Here the errors are reduced due to physical model of sensors prediction data. However, every data are from sensor nodes. The calibration between node-to-node data is utilized at a particular time on sensor. Next data is taken from chain sequentially after each data taken. It will reduce data load deployment on sensors also ensure the calibration sequential collection [21]. Multi-source data are used to estimate wireless sensor framework for air quality estimation.

3. Proposed Model

The wireless sensor networks have sensors to monitor the pollution level of the area and store it in big data server for each hour. In figure 1, the proposed architecture is explained. Hash algorithm further organizes the huge collected data and k-means clustering is done to group the similar data. Normalization in cluster is done for removing data redundancy. Air pollution analysis is done using fuzzy inference system. Ant colony optimization with genetic algorithm used for shortest and optimal path selection in Google map.

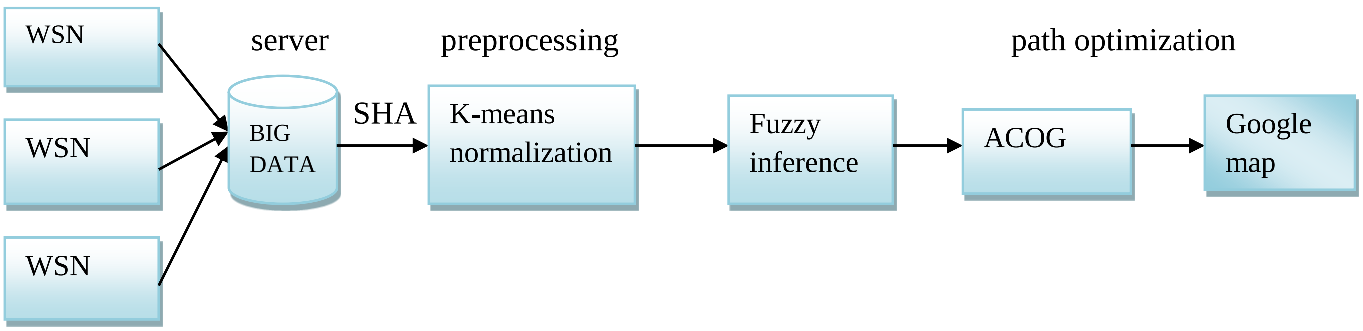

Figure 1: Big data based wireless sensor network (WSN) for path optimization

3.1 Wireless Sensor Network (WSN)

The cheapest sensor network is wireless sensor network. The sensors are placed at network nodes and connected via LAN or WAN based on the distance. Concentration on pollutants such as CO, SO_2, temperature etc are sensed and stored in the big data server. The present pollution in the air is sensed using the sensors. Sensors are simple hardware devices which measures the chemical changes in the air in its physical presence. The independent sensors are distributed across the environment to check its factors such as temperature, pressure, pollutants level. The concentration of the pollutants is checked using sensor node and transmits the data to the big data server.

The WSN have sensor nodes made up of embedded system for industrial applications. Environments of the industries are monitored using sensor nodes. Sensors are composed of temperature sensor and gas sensor components with microcontroller interface. The outputs are shown in the LC display and Zigbee system. This microcontroller is utmost admired board with several programming friendly approaches. It does not require new hardware to program and also reprogramming is very easy.

3.2 Big Data Server

The big data storage supports huge data in terabytes to store and process at a time. The data is processed using Hadoop distributed file system (HDFS). The map and reduce function are used to store the data in particular location. This mapper-reduction design used to process large information by splitting them into small process. It can process the large cluster system. Apache hadoop works with high reliability and scalability. The map reduce uses some distributed algorithm to process the unstructured datasets in hadoop environment. Map and reduce are two separate functions as given. The WSN output is transferred via WAN or LAN network to the server.

Mapper ():

The system termed as master node service is needed to perform mapper function. It splits large data into small data with some concepts. Data are scattered across the nodes and acknowledge the master node. The input work is segregated and allocated to every map node. Then the output is sending to reduce function.

Reducer ():

The segregated output is processed using reduce function. They combine the data based on some cluster techniques. They combine the spread data in to slave nodes.

3.3 PREPROCESSING

3.3.1. Hash Value Distribution

The length of the input data is 128 bit. Then the data is separated in to blocks. Constants center 8 is used to derive the 64 bit word. They are arranged as square root of first 8 prime numbers. Further process, 512 bit is updated in the buffer. In figure 2, the hash working principle is explained.

$h_{v(d_n)} = \{v(d1), v(d2) \ldots\ldots v(dn)\}$............... (1)

$h_{v(d_n)}$ is represented as hash value function for the data 'n'. V (d1) is hash value of data 1.

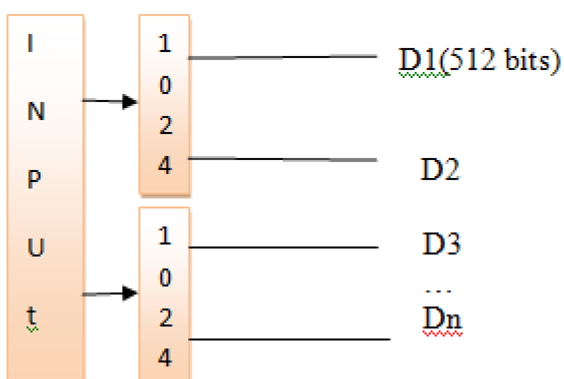

Figure 2: Hash Value Allocation

3.3.2 Feature Selection Model

1. The mostly occurred data is collected as most frequently occurred closest data as

$c_{di} = \{d1, d2 \ldots dn\}$……. (2)

2. **Support (S):** how frequently data occurs in the item set.

$data(i) \rightarrow itemset$……… (3)

3. **Confidence (C):** the total number of times data found to be true

$probability\left(\frac{item}{dataset}\right) = \frac{p(items \cup dataset)}{p(dataset)}$……… (4)

4. Probability distribution and related variance are calculated as entropy as

$entropy(S) = -\sum_{i=1}^{n} probability(S) \log_2(probability(S))$ …… (5)

$entropy(C) = -\sum_{i=1}^{n} probability(C) \log_2(probability(C))$ ……….(6)

3.3.3 k-means Clustering and Normalization

1. The minimum value and maximum value of support, confidence is taken
2. Entropy value is initialized
3. Normalization is performed for deleting repeated data as
   $N_v = \frac{data - mini(data)}{\max(data) - \min(data)}$ ……………… (7)
4. Centroids values are initialized in k-means clustering
5. Data are clustered.

Algorithm 1: k-means Normalization

Input: S, C, entropy
Output: clusters c = (c1,c2,…cn)

```
Start
      Initialize 'min' as minimum iteration and 'max' as maximum iteration
      Compute normalization
        For each N(i) do
           Find minimum distance between data and center of cluster
     End.
 While min < max
     For each data and cluster
     Initialize the new cluster center
End
For each N(i) do
           Find minimum distance between data and center of cluster
     End.
Increase min=min+1;
```

https://doi.org/10.5281/zenodo.13341530

End while
End.

The following steps explains clustered the data to its nearest centroids.

Step 1: Cluster head is represented as CH={ CH_1,$CH_{2,\ldots..}$$CH_N$}

Step 2: The every clusters are assigned to its closest CH. The distance between cluster heads and cluster data are determined using Euclidean distance

$$ED = \sqrt{(m_1 - n_1)^2 + (m_2 - n_2)^2} \ldots\ldots\ldots\ldots.. (8)$$

Where, (m_1, n_1), (m_2, n_2) are the coordinates of cluster heads and nodes. Above equation calculates the overall distance between the cluster heads and nodes in the k-means computation.

Step 3: Select the cluster node randomly such that energy of cluster might be higher than the average energy of overall cluster nodes.

Step 4: Now reassign the every cluster nodes its new cluster heads to getting best clustering result. Next, calculate the Euclidean value using equation 8, again re-evaluate the distances of all cluster nodes.

Step 5: Finally terminate the algorithm if the current distance in similar to former distance. Else, go back to Step 3.

3.3.4 Fuzzy inference framework:

This framework helps to learn and adopt the model. It has five layers to process. The clusters are fed as the input to the first layer. Layers 1 and 4 are adaptive nodes, layer 5,2 and 3 are fixed type nodes. It works as following rules.

Rule I: consider the fuzzy sets as Ai, Bi, then

$$z_i = p_i data_i + \ q_i data_{i+1} + s_i \ \ldots\ldots.. (8)$$

Rule II: if there is values, $data_i$ and $data_{i+1}$, they are features.

$$z_{i+1} = p_{i+1} data_i + q_{i+1}\, data_{i+1} + s_i \ldots\ldots\ldots\ldots\ldots\ldots (9)$$

Where, p, q, s are considered as parameter set. The output of the fuzzy is classified as good, moderate, bad, very bad, danger, satisfactory.

Google Map Air Quality Tracking

The predicted data are applied to Google map. Air quality index is listed as, If data is less than 50, then it is termed as good. If data is greater than 50 and less than 100 then it termed as moderate. If data is greater than 100 and less than 150 then it termed as satisfactory. If data is greater than 150

https://doi.org/10.5281/zenodo.13341530

and less than 200 then it termed as bad. Table 1, below shows the air quality index (AQI) readings of different area.

Agra	105
Ahmedabad	55
Aizawl	16
Ajmer	40
Alwar	60
Amaravati	37
Ambala	46
Amritsar	50
Ankleshwar	100

Table 1: value of AQI in different areas

The data shown figure 3,4 is the map view for tracking the pollution free shortest path.

Figure 3: AQI in Google map

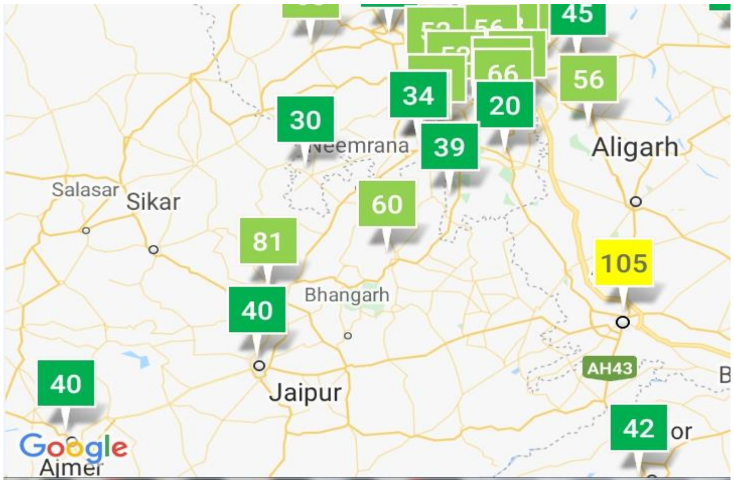

Figure 4: AQI in Delhi, India

5. ACO and Genetic Algorithm(GA) Process

5.1 Ant Colony Optimization

Its main objective is to tackle the travelling sales man problem. Every ant follows previous route of the ant to search the food in easiest way. This is achieved ant's pheromone value which is left by every ant. Its target is to connect various nodes with their shortest path. The full solution can be obtained by including components in every iteration to solve feasible solution. ACO does not perform backtracking. It's is an alternative process for greedy algorithm. The ants searching for their food is the inspiration for this ACO algorithm. Pheromone value is left by ant during their search, which is path hint for other ants. The algorithm uses numerical value to represents pheromone information. ACO uses discrete principal in optimization problem.

To increase the quality of the optimization, the hybrid approach is used by combining ACO with genetic and local search algorithm. Local search improves the solution by looking into neighborhood solution each time. During searching operation, it produces best results.

Genetic algorithm computes for initializing the database (db). It considers three components as, labels for every variable, member function from scaling of nonlinear function, scaling parameters. The chromosome population is used to initialize the genetic process. First generation evaluation mentioned as G(0). Next evaluation function continues till it meets the termination rule. New population in the genetic algorithm is evaluated as (G(P+1)) from every individual genetic chromosome. Further detailed study can be seen in [31].

5.2 Steps to Process ACO with Genetic Algorithm

1. Genetic algorithm DB is generated
2. Use the fuzzy rules initiated in equation 8 and 9.
3. Select the rule for every subspace using two steps.

   a. for every subspace (S) , generate the required data consequents (c_i) as, {S= c1,c2,……cn}, N_S is the total number of subspaces.

b. Run ACO algorithm 2 for best accuracy and throughput.

4. Randomly generated possible solutions are initialized.

Randomly chosen subspaces are belongs to {S= c1,c2,……cn},. And changes the data Consequents (c_i)

Quality of solution is based on set on rule encoded. Mean square value (MSV) is calculated as,

$$MSV = \frac{1}{2N}\sum_{i=1}^{n}\left(X^i - x^i\right)^2 \text{......................(10)}$$

The 'X' is a obtained output from the 'x' desired output. If the outcome is near to zero, then it is termed as global optimal performance.

5. The fitness value is calculated as , $fitness = \omega_{i.}MSV + \omega_{i.}rule$ …… (11)

6. End the process if condition meets, else go to step 1.

5.3 Searching the Pollution Free Optimal Path

Initially ant starts from departure place to reach destination. Decision is made from pheromone level from previous ant. Here pheromone level is probability of choosing the route. Ant move from one node to another node with pheromone level and less pollution level. The β_{ij} factor used to handle heuristic data. This factor helps to look for past move and compute present moves. Ants starts at initial node and completes at final node. Movement of the ant is nothing but movement of car from one node to another node.

Lets consider the mathematical computation for ant 'x' moving from node a to b, then S_{ab}^{X} represents the model.

$$S_{ab}^{X} = \begin{cases} \frac{p_{ab}^{\alpha}.h_{ab}^{\beta}}{\sum_{y \epsilon Q_a^x} p_{ay}^{\alpha}.h_{ay}^{\beta}} \; if \; i \epsilon Q_a^x \\ 0 \qquad otherwise \end{cases} \text{.....................(12)}$$

In equation 12 , s nodes with ant 'x' visit node 'a', where s={a,b.c…..etc). h_{ab}^{β} represents, heuristic details of nodes a and b. 'p' represents pheromone value. The pheromone value will be updated when cars (ants) finish their tour. Value updation is done by reducing the constant factor to prevent algorithm to take bad decision and set to limited pheromone implementation.

Updated value p_{ab}^{α} is kept on arc between the nodes, which ants have crossed its path recently with evaporation rate in equation 13 and pheromone deposition as equation 14.

$p_{ab} = (1-\gamma)p_{ab}$ ….. (13) ,where γ lies between 0 to 1 as evaporation rate.

$p_{ab} = p_{ab} + \sum_{x=1}^{n} \Delta \; S_{ab}^{X}$ ……… (14)

$$\Delta S_{ab}^{X} = \begin{cases} \frac{1}{C^K}, & if\ arc\ from\ a\ to\ b\ reprsents\ as\ G^K \\ 0, & else \end{cases} \quad \ldots\ldots (15)$$

C^K represents cost of the arc. It is to be noted genetic step, crossover, mutations takes first before pheromone trails updates its value. In our work two-point crossover is included as follows:

$v_1 = \frac{|s_{ab}|}{3.0}$ …………..(16)
$v_2 = v_1 + \frac{|s_{ab}|}{2}$…………(17)

Hence v1 and v2 are the crossover points of nodes and then mutation process is performed by replacing the new genes in chromosomes. The gene replaced are new to chromosome, it is not repeated. Chromosomes are the solution finally.

The value of pheromone is strengthened using following equations,

$p_{ab} = p_{ab} + \sum_{x=1}^{n} \Delta\ S_{ab}^{X} + \Delta s_{ab}^{best}$……(18)

$$\Delta s_{ab}^{best} = \begin{cases} \frac{1}{C^{best}}, & if\ arc\ from\ a\ to\ b\ reprsents\ as\ G^{best} \\ 0, & else \end{cases} \quad \ldots\ldots\ldots..(19)$$

Equation 19explains the best pheromone value still now.

Algorithm 2: ACOG

Begin
Initialize ([number clusters], [number ants]);
Repeat
1. For each ant do M:
2. For each node Calc(probability belonging node to cluster) ;
3. If node=moderate||good
4. Select and set pheromone value.
5. Node =node+1;
6. Step 3 to 5.
7. If node= destination stop

End

8. Update (cluster center);
9. If (NewCenter < > OldCenter)
then goto M;

Else
Save(current solution);
End
select Best Solution From All Ants ;
Update (for each node) ;
Correct (common solution);

Anadolu Psikiyatri Dergisi || ISSN: 1302-6631 || Volume 25; Number 4
https://doi.org/10.5281/zenodo.13341530

Until criteria not reached .
End

The algorithm computes the best travelling path from the starting node to destination node. The optimal solution is selected from all the available solutions.

Result and Evaluation

The performance of every node is evaluated on its time basis using the java platform. The time of proposed algorithm to reach the destination via pollution free route is very less when compared to existing algorithm.

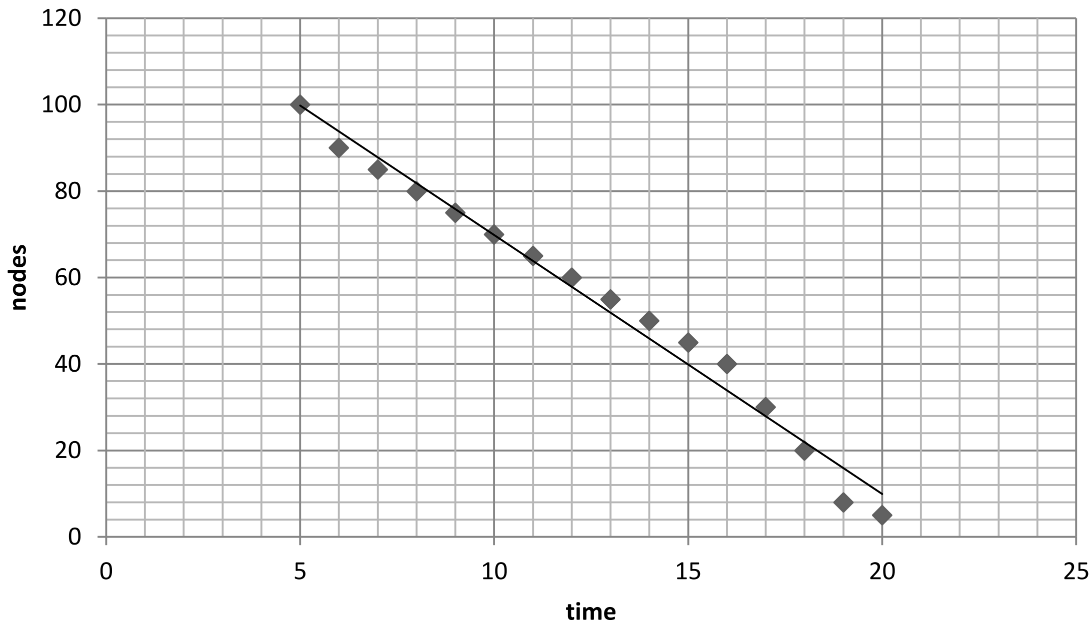

Figure 5: ACGO Time Tracking

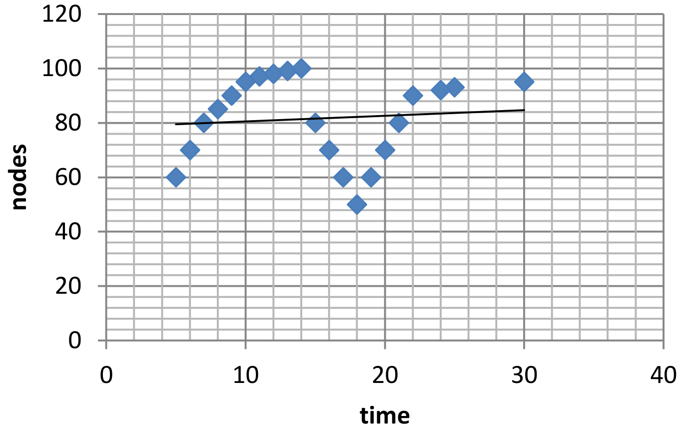

Figure 6: Krill Herd Time Tracking

From the above graph in figure 5 and 6, its noted that the destination is reached in 20 min using ACOG where as krill herding takes 30 minutes to reach the same destination.

Accuracy of the ACOG algorithm is compared with krill herd algorithm, genetic algorithm. ACOG gives better accuracy and throughput of selecting the global optimal path in the system. The below figure 7,8 shows the accuracy and throughput comparison.

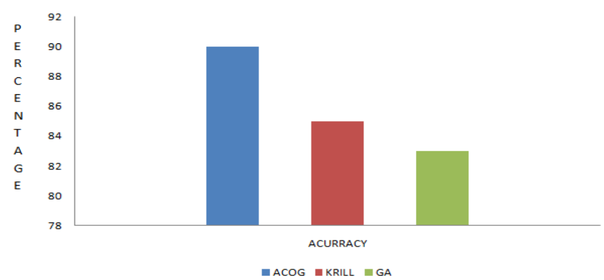

Figure 7: Accuracy Comparison Graph

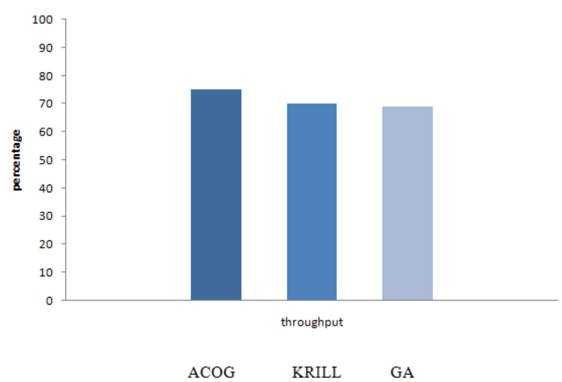

Figure 8: Through Put Comparison Graph

Conclusion and Future Scope

The health of every human being is affected by some airborne diseases. Pollution is heavy due to industrial developments. Mining industries, manufacturing industries, lakh of vehicle movement causes severe air pollution in surrounding environment. Lung cancer, acid rains, viral infections, breathing troubles is some issues faced by present generation. This problem is addressed by recent technological development with wireless sensor networks and big data handling huge network data. The ant colony optimization is most widely used optimization technique in swarm intelligence platform. Its saves the time of other ants to search route for the food. In our proposed work, we sense the severity of the air pollution and based on the air quality index ACOG algorithm tracks the shortest route in the Google map. Healthy initiative travel can be achieved by our proposed work. It achieves higher accuracy when compared with existing systems. In future it can be introduce various hybrid approaches to air pollution problems. There are numerous high level swarm intelligence technique are introduced in recent days.

References

[1] M. Tang, X. Wu, P. Agrawal, S. Pongpaichet, and R. Jain, ``Integration of diverse data sources for spatial PM2.5 data interpolation,'' IEEE Trans. Multimedia, vol. 19, no. 2, pp. 408_417, Feb. 2017. doi: 10.1109/TMM. 2016.2613639.

[2] U. K. von Schwedler, M. Stuchell, B. Müller, D. M. Ward, H.-Y. Chung, E. Morita, H. E. Wang, T. Davis, G.-P. He, D. M. Cimbora, A. Scott, H.-G. Kräusslich, J. Kaplan, S. G. Morham, andW. I. Sundquist, ``The protein network of HIV budding,'' Cell, vol. 114, no. 6, pp. 701_713, Oct. 2003. doi: 10.1016/s0092-8674(03)00714-1.

[3] C. J. Fang, F. J. Shao, Y. Sui and R. C. Sun, ``Analysis of meteorological network based on complex network,'' J. Qingdao Univ., vol. 30, no. 4, pp. 50_57, Nov. 2017.

[4] M. A. Al-Garadi, K. D. Varathan, S. D. Ravana, E. Ahmed, G. Mujtaba, M. U. S. Khan, and S. U. Khan, ``Analysis of online social network connections for identification of influential users: Survey and open research issues,'' ACM Comput. Surv., vol. 51, no. 1, p. 16, Jan. 2018. doi: 10.1145/ 3155897.

[5] F. M. Lopes, R. M. Cesar, Jr., and L. D. F. Costa, ``Gene expression complex networks: Synthesis, identification, and analysis,'' J. Comput. Biol., vol. 18, no. 10, pp. 1353_1367, May 2011. doi: 10.1089/cmb.2010.0118.

[6] M. Kampa and E. Castanas, ``Human health effects of air pollution,'' Environ. Pollut., vol. 151, no. 2, pp. 362_367, Jan. 2008.

[7] E. Boldo, S. Medina, A. Le Tertre, F. Hurley, H.-G. Mücke, F. Ballester, and I. Aguilera, ``Apheis: Health impact assessment of long-term exposure to PM2.5 in 23 European cities,'' Eur. J. Epidemiology, vol. 21, no. 6, pp. 449_458, Jun. 2006.

[8] J. Lin, A. Zhang, W. Chen, and M. Lin, ``Estimates of daily PM2.5 exposure in beijing using spatio-temporal kriging model,'' Sustainability, vol. 10, no. 8, p. 2772, 2018.

[9] Y. Jiang, L. Shang, K. Li, L. Tian, R. Piedrahita, X. Yun, O. Mansata, Q. Lv, R. P. Dick, and M. Hannigan, ``MAQS: A personalized mobile sensing system for indoor air quality monitoring," in Proc. 13th Int. Conf. Ubiquitous Comput. UbiComp, 2011, pp. 271_280.

[10] J. Mitrovics, ``Smart sensors for air quality monitoring: Concepts and new developments," in Proc. IEEE Sensors, Oct./Nov. 2017, pp. 1_2.

[11] Dan Zhang and Simon S. Woo, "Real Time Localized Air Quality Monitoring and Prediction Through Mobile and Fixed IoT Sensing Network" May 22, 2020. DOI: 10.1109/ACCESS.2020.2993547.

[12] Yan Kuang David He Serap Erdal Donna Kenskie Ming Dong, Dong Yang. Pm2.5 concentration prediction using hidden semi-markov model-based times series data mining. 36(5):9046–9055, 2009.

[13] Jingyi Sun, Wei; Sun. Daily pm2.5 concentration prediction based on principal component analysis and lssvm optimized by cuckoo search algorithm. 03 2017.

[14] Shan Zhang, Xiaoli Li, Yang Li, and Jianxiang Mei. Prediction of urban pm 2.5 concentration based on wavelet neural network. In 2018 Chinese Control And Decision Conference (CCDC), pages 5514–5519. IEEE, 2018.

[15] Zhigen Shang, Tong Deng, Jianqiang He, and Xiaohui Duan. A novel model for hourly PM2.5 concentration prediction based on cart and eelm. Science of The Total Environment, 651:3043–3052, 2019.

[16] Victor OK Li, Jacqueline CK Lam, Yun Chen, and Jiatao Gu. " Deep learning model to estimate air pollution using m-bp to fill in missing proxy urban data". In GLOBECOM 2017-2017 IEEE Global Communications Conference, pages 1–6.

[17] Yu Pengfei, He Juanjuan, Liu Xiaoming, and Zhang Kai. Industrial air pollution prediction using deep neural network. In *International Conference on Bio-Inspired Computing: Theories and Applications*, pages 173–185. Springer, 2018.

[18] I˙brahim Ko¨k, Mehmet Ulvi S¸ ims¸ ek, and Suat O¨ zdemir. A deep learning model for air quality prediction in smart cities. In Big Data (Big Data), 2017 IEEE International Conference on, pages 1983–1990. IEEE, 2017.

[19] Zhongang Qi, Tianchun Wang, Guojie Song, Weisong Hu, Xi Li, and Zhongfei Mark Zhang." Deep air learning: Interpolation, prediction, and feature analysis of fine-grained air quality ". IEEE Transactions on Knowledge and Data Engineering, 2018.

[20] Ahmed Boubrima, Walid Bechkit, Herv´e Rivano, and Lionel Soulhac. Poster: Toward a better monitoring of air pollution using mobile wireless sensor networks. In Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking, pages 534–536. ACM, 2017.

[21] Cheng Feng, Wendong Wang, Ye Tian, Xirong Que, and Xiangyang Gong. Air quality estimation based on multi-source heterogeneous data from wireless sensor networks. In Wireless Communications and Networking Conference (WCNC), 2018 IEEE, pages 1–6. IEEE, 2018.

[22] Movva Pavani, and P. Trinatha Rao, "Monitoring real-time urban carbon monoxide (CO) emissions using wireless sensor networks", In International Conference on Information and Communication Technology for Intelligent Systems. Springer, Cham, pp. 290-297, 2017.

[23] Byron Guanochanga, Rolando Cachipuendo, Walter Fuertes, Santiago Salvador, Diego S. Benítez, Theofilos Toulkeridis, Jenny Torres, César Villacís, Freddy Tapia, and Fausto Meneses, [24] "Real-time air pollution monitoring systems using wireless sensor networks connected in a cloud-computing, wrapped up web services", In Proceedings of the Future Technologies Conference. Springer, Cham, pp. 171-184, 2018.

[25] Amjed Belkhiri, Walid Bechkit, Hervé Rivano, and Mouloud Koudil, "Context aware MWSN optimal redeployment strategies for air pollution timely monitoring", In IEEE ICC 2018-IEEE International Conference on Communications. IEEE, 2018, pp. 1-7.

[26] Liqun Bao, and Haibin Wang, "Distributed real-time monitoring system for atmospheric particles", IET Wireless Sensor Systems, Vol. 7, No. 4, pp. 91-97, 2017.

[27] Sherin Abraham, and Xinrong Li, "Design of a low-cost wireless indoor air quality sensor network system", International Journal of Wireless Information Networks, Vol. 23, No. 1, pp. 57-65, 2016.

[28] Byron Guanochanga, Rolando Cachipuendo, Walter Fuertes, Diego S. Benítez, Theofilos Toulkeridis, Jenny Torres, César Villacís, Freddy Tapia, and Fausto Meneses, "Towards a real-time air pollution monitoring systems implemented using wireless sensor networks: preliminary results", In 2018 IEEE Colombian Conference on Communications and Computing (COLCOM), IEEE, pp. 1-4, 2018.

[29] Deepika Patil, T. C. Thanuja, and B. C. Melinamath, "Air pollution monitoring system using Wireless Sensor Network (WSN) ", In Data Management, Analytics and Innovation. Springer, Singapore, pp. 391-400, 2019.

[30] Qilong Han , Peng Liu , Haitao Zhang, and Zhipeng Cai , "A Wireless Sensor Network for Monitoring Environmental Quality in the Manufacturing Industry, 0.1109/ACCESS.2019.2920838, volume 7.

[31] O. Cordon, F. Herrera, L. Magdalena, and P. Villar, "A genetic learning process for the scaling factors, granularity and contexts of the fuzzy rule-based system data base," Information Sciences, vol. 136, pp. 85-107, 2001.