

Two Hybrid Expert System for Diagnosis Air Quality Index (AQI)

Leila Abdolkarimzadeh^(✉), Milad Azadpour, and M.H. Fazel Zarandi

Department of Industrial Engineering,
Amirkabir University of Technology, Tehran, Iran
{abdolkarimzadeh,zarandi}@aut.ac.ir

Abstract. Air pollution is a common problem in areas with high population density such as big cities. The mega city of Tehran which is capital city of Iran is suffering from poor air quality. In Tehran, Traditional air quality assessment is realized using air quality indices which are determined as max values of selected air pollutants which is mostly base on PM2.5. Thus, air quality assessment depends on strictly describe without taking into account specific other Environmental parameters. In this paper, To demonstrate the application, common air pollutants like CO, O3, NO2, SO2, PM10 and PM2.5 are used as air pollutant parameters, also we were studied over an 2-year period (2015–2017) on daily data of the air quality index (AQI) in Tehran. The artificial intelligence based on neural network and fuzzy inferences methods allows assessing air quality parameters, providing a partial solution to this problem. Accordingly, this study proposes two fuzzy logic system for assessing accurate air quality evaluations, also proposed Seven score stages: Good, Moderate, Unhealthy for Sensitive Group, Unhealthy, Very Unhealthy, Hazardous for evaluating air quality index. Our experimental results show a good performance of the proposed air quality index against other system that those in literature.

Keywords: AQI · Air pollution · Expert system

1 Introduction

Nowadays air pollution has become a crisis in people's daily lives and had many adverse effects on human health, so when air quality is poor in terms of the number of people referred to treatment centers significantly increased. For people with chronic bronchitis, emphysema, asthma high concentrations of pollution can cause breathing difficulties. In addition, for older people with heart problems and respiratory diseases, increasing pollutants particles levels can cause premature death. Therefore, pollutants such as CO, O3, NO2, SO2, PM10 are major factors that affected on Air Quality Index. PM10 is one of the major factors contributing to problems caused by air pollution [1]; another factors are NO2, CO, O3, SO2 that the damage of each pollutant causes separately is known [2] thus, it appears

that obtaining measurements of air pollutants based on all factors observations in urbanized regions is essential. In Tehran approximately 600 thousands cars are used daily, According to recent research percentage of emissions matter such as NO₂, SO₂, CO₂, CO 64.3, 29.3, 27.5, 98.6% respectively. And 79.2% of hovering particles is produced in country Originate from the Transportation Industry, whereas, cars have a significant share of the total pollution So that in Tehran 52% of CO and 41% of hovering particles (particulate matter) comes these cars. In 2010, CO was the most important pollutants, after that in the next rank O₃ and PM were most polluted days of the year. The Rest of this research is organized as Sect. 2 explain the Background pollutants and their main characteristics in air quality assessment. In Sect. 3, our method that is a Fuzzy Inference System for air quality assessment is proposed, also this section shows the performance and efficiency of our system. In Sect. 4, describe and show result of the validity of the fuzzy system. Finally, Sect. 5 reported the conclusion of the study.

2 Background

According to this problem, international organizations have implemented similar methodologies for air pollutant assessment and monitoring such as the Environmental Protection Agency in United States [2] and the Pan American Health Organization [3]. Human permissible limits, are in a ranges set and are used to calculate the AQI index in Iran. Artificial neural networks [4]; support vector machines [5], amongst others. Alternatively, other methodologies have been proposed for evaluating air quality using computational models such as fuzzy logic [6]; Mishra and Goyal [7]. Olvera-García et al. [1] air quality assessment using fuzzy logic and combining an Analytic Hierarchy Process (AHP) have been developed, providing different solutions, Upadhyay et al. [8], Akkaya et al. 2015 [8]. In the present study, the proposed models are applied for analyzing the air quality Index of Tehran City and proposed a fuzzy logic system base on TSK. In Tehran Some area has special indicator for diagnose Air Quality Index about 40 station that send data to central station. In Fig. 1, we indicate some station in Iran like Region 2, Shadabad, Fath, Piroozi, respectively.

2.1 AQI Index

The United States Environmental Protection Agency (EPA) has developed an Air Quality Index that is used to report air quality. This AQI is divided into six categories indicating increasing levels of health concern. The EPA has established National Ambient Air Quality Standards (NAAQS) for each of these pollutants in order to protect public health. Tables 1 and 2 indicate pollutants and their health impacts Based on the measured ambient concentrations.

2.2 Case Study

Iran is located in the Middle East, between Iraq and Pakistan bordering the Gulf of Oman, the Persian Gulf, and the Caspian Sea that is the Coastline of Iran

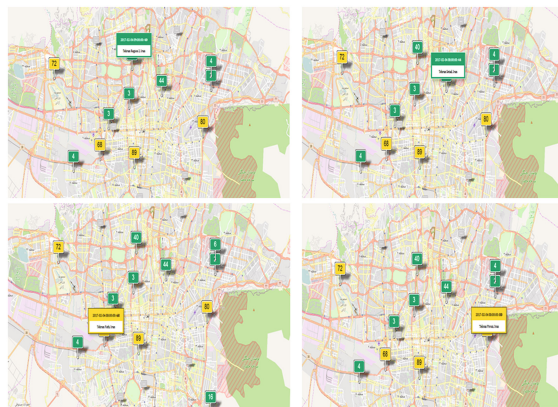


Fig. 1. Some stations measuring air pollution in Tehran

Table 1. The range of pollutants

AQI category (Range)	CO (8 h)	O3 (8 h)	NO2 (1 h)	SO2 (24 h)	PM10 (24 h)	PM2.5 (24 h)
Good	0–4.4	0–0.059	0–0.053	0–0.034	0–54	0–15.4
Moderate	4.5–9.4	0.060–0.075	0.054–0.1	0.035–0.144	55–154	15.5–35
Unhealthy for sensitive groups	9.5–12.4	0.076–0.095	0.101–0.360	0.145–0.224	155–254	35.1–65.4
Unhealthy	12.5–15.4	0.096–0.115	0.361–0.640	0.225–0.304	255–354	65.5–150.4
Very unhealthy	15.5–30.4	0.116–0.374	0.65–1.24	0.305–0.604	355–424	150.5–250.4
Hazardous	30.5–40.4	0.405–0.504 (1 h)	1.25–1.64	0.605–0.804	425–504	250.5–350.4
	40.5–50.4	0.505–0.604 (1 h)	1.65–2.04	0.805–1.004	505–604	350.5–500.4

Table 2. categories of AQI index

AQI	Associated health impacts
Good (0–50)	Minimal impact
Moderate (51–100)	May cause minor breathing discomfort to sensitive people
Unhealthy for sensitive groups (101–150)	May cause breathing discomfort to people with lung disease such as asthma, and discomfort to people with heart disease, children and older adults
Unhealthy (151–200)	May cause breathing discomfort to people on prolonged exposure, and discomfort to people with heart disease
Very unhealthy (201–300)	May cause respiratory illness to the people on prolonged exposure. Effect may be more pronounced in people with lung and heart diseases
Hazardous (301–400), (401–500)	May cause respiratory impact even on healthy people, and serious health impacts on people with lung/heart disease. The health impacts may be experienced even during light physical activity

is 2,440 km. Tehran is the capital city of Iran and is located at the foot of the towering Alborz mountain range also, a bustling metropolis of 14 million people, it's the largest urban city in Western Asia. The common environmental threats to the geography of Iran is air pollution. In this study, air quality monitoring data for different areas of Tehran. The issue is very sensitive to air pollution and the preference is to be taken advantage of all available data. The system architecture used C-means clustering [9] and definition fuzzy inference system desired. In Fig. 2 illustrate our variables which should include CO, O3, NO2, SO2, PM10, PM2.5.

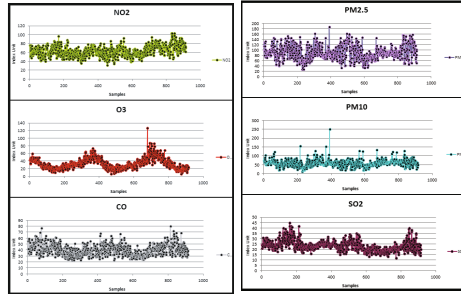


Fig. 2. Data analysis

3 Method

3.1 The Structure of Fuzzy System

To deal with monitoring air quality, especially in urban areas has become an essential and vital. For accurate monitoring of air quality standards are often defined by international and domestic institutions and to observe any contaminants of health determined, Most methods of measuring air pollution cites the recommendations of these institutions. To determine the level of contamination is defined as the extent to which any emissions As well as qualitative terms are defined linguistic variables to define air pollution Are ambiguities that sometimes change a small amount of a pollutant, air pollution is on another level interpretation. This uncertainty and confusion in the field of air pollution monitoring in the form of fuzzy systems, fuzzy logic can apply to a large extent be covered. Fuzzy inference system with fuzzy input variables in the input certainty with ambiguity and error is rejected And by combining inputs and outputs using fuzzy inference that is more reliable Because a large extent eliminates the uncertainty and ambiguity of input values and their effect on the final result looked more closely. In this study two fuzzy system is designed to monitor air quality based on certain concepts to classify air quality at different times. Both systems use the data to monitor the amount of pollutants contributing to air pollution in Tehran (winter 2015 to winter 2017) are designed. The first simulation system for AQI standard behavior data used in calculating the level of air pollution in Tehran. The main purpose of this study is determining the level of air quality present based on a number of language variables, so we are facing a problem of categorization. These variables are as following. **Good, Moderate, Unhealthy for Sensitive Groups, Unhealthy, Very Unhealthy, Hazardous**, also in the design of their systems of 6 input variables that we use the same contaminants. Fuzzy input variables which should include **CO, O3, NO2, SO2, PM10, PM2.5** are For fuzzy inference in this study have chosen **Sugeno Fuzzy Inference System**. In previous studies used Mamdani fuzzy system for determining rules, Defuzzification and direct separation of data. We use Sugeno fuzzy inference system that show in Fig. 3.

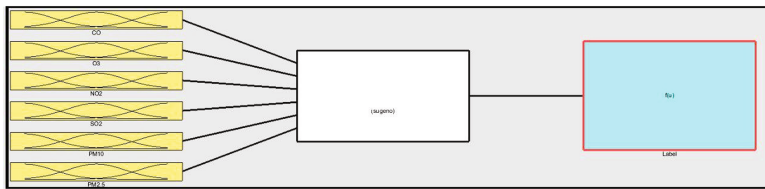


Fig. 3. TSK system

3.2 Fuzzy Rules

Fuzzy rules in the fuzzy inference systems are used for rule if - then defined. Various methods for extracting rule. In both systems designed fuzzy clustering method to extract rules (subtractive clustering). For clustering method of fuzzy is used Sugeno fuzzy rules in the system as defined in Fig. 4. In this phase systems are faced with MISO rules as to the number of air pollutants that are 6 to input variable. And only the output of our system is the same label. Output Sugeno fuzzy rules in relationships based on input variables that must be set coefficients. Classification issues can be considered zero coefficients input variables and the output of each rule is equal to the label that it displays.

Rule 1: IF CO isr $MF_{CO,1}$ AND O_3 isr $MF_{O_3,1}$ AND NO_2 isr $MF_{NO_2,1}$ AND SO_2 isr $MF_{SO_2,1}$ AND PM_{10} isr $MF_{PM_{10,1}}$ AND $PM_{2.5}$ isr $MF_{PM_{2.5,1}}$ THEN AQIC isr $F_{11}(u)$.

Rule 2: IF CO isr $MF_{CO,2}$ AND O_3 isr $MF_{O_3,2}$ AND NO_2 isr $MF_{NO_2,2}$ AND SO_2 isr $MF_{SO_2,2}$ AND PM_{10} isr $MF_{PM_{10,2}}$ AND $PM_{2.5}$ isr $MF_{PM_{2.5,2}}$ THEN AQIC isr $F_{12}(u)$.

Rule 3: IF CO isr $MF_{CO,3}$ AND O_3 isr $MF_{O_3,3}$ AND NO_2 isr $MF_{NO_2,3}$ AND SO_2 isr $MF_{SO_2,3}$ AND PM_{10} isr $MF_{PM_{10,3}}$ AND $PM_{2.5}$ isr $MF_{PM_{2.5,3}}$ THEN AQIC isr $F_{13}(u)$.

Rule 4: IF CO isr $MF_{CO,4}$ AND O_3 isr $MF_{O_3,4}$ AND NO_2 isr $MF_{NO_2,4}$ AND SO_2 isr $MF_{SO_2,4}$ AND PM_{10} isr $MF_{PM_{10,4}}$ AND $PM_{2.5}$ isr $MF_{PM_{2.5,4}}$ THEN AQIC isr $F_{14}(u)$.

Rule 5: IF CO isr $MF_{CO,5}$ AND O_3 isr $MF_{O_3,5}$ AND NO_2 isr $MF_{NO_2,5}$ AND SO_2 isr $MF_{SO_2,5}$ AND PM_{10} isr $MF_{PM_{10,5}}$ AND $PM_{2.5}$ isr $MF_{PM_{2.5,5}}$ THEN AQIC isr $F_{15}(u)$.

Rule 6: IF CO isr $MF_{CO,6}$ AND O_3 isr $MF_{O_3,6}$ AND NO_2 isr $MF_{NO_2,6}$ AND SO_2 isr $MF_{SO_2,6}$ AND PM_{10} isr $MF_{PM_{10,6}}$ AND $PM_{2.5}$ isr $MF_{PM_{2.5,6}}$ THEN AQIC isr $F_{16}(u)$.

In this study, fuzzy sets has been considered according to final data Labels. Six labels must be calculated in the series for each variable as fuzzy set membership function. Linguistic variables synonymous with the six sets are as follows: Very low, low, medium, high, very high, bulky, corresponding to the labels are output.

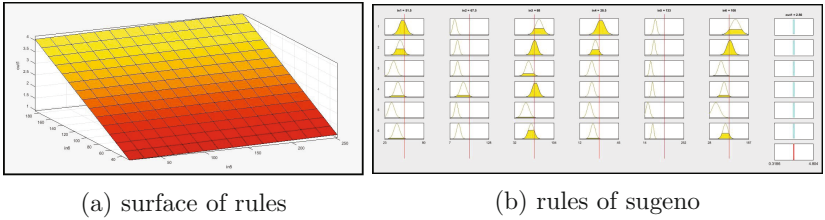


Fig. 4. Rules of sugeno

3.3 Membership Functions

In type 1 for each variable fuzzy logic membership functions defined fuzzy membership value taken by the variable (in the system at any given moment) in the fuzzy set of pre-defined number between 0 and 1 to indicate In fuzzy logic membership functions can have different states depending on the nature of the problem and its parameters must be set. In a fuzzy system and fuzzy variable is considered that the amount of these pollutants. This level of pollutants Unlike previous systems by assigning different values for member states fuzzy sets. With this approach, there may be a certain amount of pollutants to also have a large collection of membership And in the middle set, but with a different membership levels and complement each other, this approach increases the accuracy and transparency of future decisions. In the first system of Gaussian membership functions we have used to define fuzzy numbers Gaussian membership function has two parameters, the mean and variance Experience has shown that the variables With the continuous high degree of this type of membership functions provide better coverage of uncertainty Since we have 6 input variables and each variable in each of the rules are to define 36 different membership function; To do this it is necessary that we set the parameter 72. In the case of setting the parameters of membership functions of the system do the work. Once we set the parameters using ANFIS system that show in Fig. 5, and gain accuracy of the system. Using Genetic Algorithm optimization parameters of membership functions to configure again and we calculate the final accuracy of the system. Finally, each of the sets of parameters that have to provide more accuracy Picking

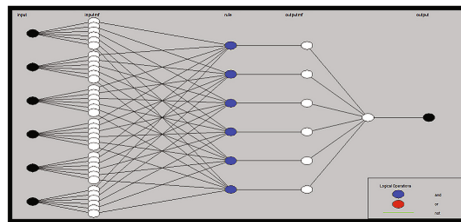
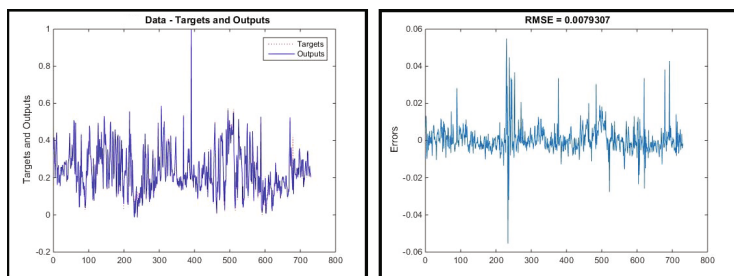


Fig. 5. ANFIS

out. In the second system to determine the membership functions according to PSO algorithm. Using PSO Algorithm optimization parameters of membership functions to configure again and we calculate the final accuracy of the system. The second system diagnosis AQI index better than first system. Since these two systems have been designed using Sugeno we do not need Defuzzification, Because the output Sugeno rules as well as the entire system is an absolute number to each variable output.

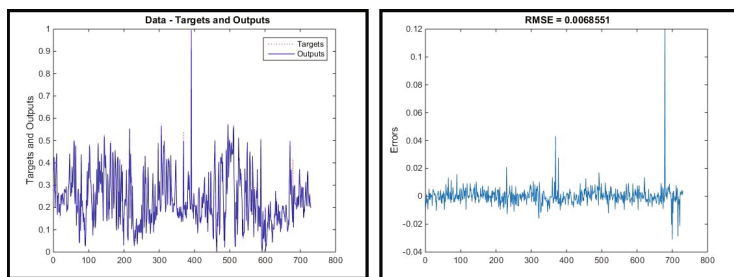
4 Validation System

Designing Sugeno fuzzy system by Fuzzy c-means clustering, also use Neural network for setting parameters of sugeno system for improving performance of the system. In during training parameters with ANFIS and GA, and the second system improved parameters with PSO. Accuracy of the output's system is increased. Figures 6 and 7 illustrate, the comparison between AQI and output of the two system that is designed for training and testing data. According our



(a) The result of AQI and output by ANFIS with GA (b) Error of training data by ANFIS with GA

Fig. 6. First system



(a) The result of AQI and output by ANFIS with PSO (b) Error of training data by ANFIS with PSO

Fig. 7. Second system

analysis, system could create suitable AQI on real output, although error of the system is reduced.

5 Conclusion

In this study, Two new model for air quality assessment has been built as a way to evaluate Air Quality Index in urban cities. First model was created according to the environment dynamic located in the Tehran City, where the high density population. In order to generate more complete air quality analyses, six major pollutants were measured, AQI-ANFIS with GA provided a good integration all of particular evaluations; however, the priority assessment is focused on the air quality assessment to those parameters that represent the major problem on human health. The AQI-ANFIS with PSO can be adjusted to define other parameters that represent critical problems in urban areas. An analysis about a fitted membership function type for a Sugeno (TSK), specific parameter is tuned. Additionally, a knowledge base which introduces new rules based on parameter behaviors is a good idea for having an improved computational model.

References

1. Olvera-García, M.Á., Carbajal-Hernández, J.J., Sánchez-Fernández, L.P., Hernández-Bautista, I.: Air quality assessment using a weighted fuzzy inference system. *Ecol. Inf.* **33**, 57–74 (2016)
2. Carbajal-Hernández, J.J., Sánchez-Fernández, L.P., Carrasco-Ochoa, J.A., Martínez-Trinidad, J.F.: Assessment and prediction of air quality using fuzzy logic and autoregressive models. *Atmos. Environ.* **60**, 37–50 (2012)
3. Aquino, R., de Oliveira, N.F., Barreto, M.L.: Impact of the family health program on infant mortality in Brazilian municipalities. *Am. J. Public Health* **99**(1), 87–93 (2009)
4. Salazar-Ruiz, E., Ordieres, J., Vergara, E., Capuz-Rizo, S.F.: Development and comparative analysis of tropospheric ozone prediction models using linear and artificial intelligence-based models in Mexicali, Baja California (Mexico) and Calexico, California (US). *Environmental Modelling & Software* **23**(8), 1056–1069 (2008)
5. Wang, W., Men, C., Lu, W.: Online prediction model based on support vector machine. *Neurocomputing* **71**(4), 550–558 (2008)
6. Liu, K.F., Liang, H., Yeh, K., Chen, C.: A qualitative decision support for environmental impact assessment using fuzzy logic. *J. Environ. Inform.* **13**(2), 93–103 (2009)
7. Mishra, D., Goyal, P.: Neuro-fuzzy approach to forecast NO₂ pollutants addressed to air quality dispersion model over Delhi, India. *Aerosol Air Qual. Res.* **16**, 166–174 (2016)
8. Upadhyay, A., Kanchan, P.G., Yerramilli, A., Gorai, A.K.: Development of fuzzy pattern recognition model for air quality assessment of howrah city. *Aerosol Air Qual. Res.* **14**, 1639–1652 (2014)
9. Zarandi, M.F., Kalhori, M.R.N., Jahromi, M.: Possibilistic c-means clustering using fuzzy relations. In: 2013 Joint IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS), pp. 1137–1142. IEEE (2013)