

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/250219902>

Application of fuzzy c-means clustering technique in vehicular pollution

Article · July 2013

DOI: 10.5937/JPMNT1303094D

CITATIONS

7

READS

162

2 authors, including:



[Hemanta K. Baruah](#)

Bodoland University

116 PUBLICATIONS 527 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



I am currently thinking about how to use my theory of regions to solve graph theoretic problems. I had used my theory of regions to prove the four color theorem mathematically a few years back. [View project](#)

All content following this page was uploaded by [Hemanta K. Baruah](#) on 10 September 2017.

The user has requested enhancement of the downloaded file.

APPLICATION OF FUZZY C-MEANS CLUSTERING TECHNIQUE IN VEHICULAR POLLUTION

Samarjit Das¹ and Hemanta K. Baruah²

¹Department of Computer Science & IT, Cotton College, Assam, India

²Department of Statistics, Gauhati University, Assam, India

¹ssainm@rediffmail.com, ²hemanta_bh@yahoo.com

Abstract- Presently in most of the urban areas all over the world, due to the exponential increase in traffic, vehicular pollution has become one of the key contributors to air pollution. As uncertainty prevails in the process of designating the level of pollution of a particular region, a fuzzy method can be applied to see the membership values of that region to a number of predefined clusters. Also, due to the existence of different pollutants in vehicular pollution, the data used to represent it are in the form of numerical vectors. In our work, we shall apply the fuzzy c-means technique of Bezdek on a dataset representing vehicular pollution to obtain the membership values of pollution due to vehicular emission of a city to one or more of some predefined clusters. We shall try also to see the benefits of adopting a fuzzy approach over the traditional way of determining the level of pollution of the particular region.

Keywords: Vehicular pollution, fuzzy c-means technique.

1. Introduction

Air pollution has been aggravated due to the rapid growth of cities, increasing traffic, industrialization and higher levels of energy consumption. Currently, all over the world, air pollution is widespread in urban areas where vehicles are the major contributors and in some other areas with high concentration of industries and thermal power plants. Vehicular emissions are of particular concern since these are ground level sources and thus have the maximum impact on the general population. Motor vehicles produce harmful air emissions which have various adverse effects not only on human beings but also

on the entire environment. Therefore it is very essential to make the common people aware of the present status of pollution along with its key causes so as to prevent it from creating dreadful conditions. Various government and non-government organizations have already taken initiative in this regard. From the literature, it can be seen that the whole range of pollution has been divided into some categories which can also be considered as clusters, for example from 'Good' to 'Hazardous' and the level of pollution of a particular region or city is bound to belong to exactly one of the predefined clusters. In the present work, we shall try to see the level of pollution of a particular city due to vehicular emissions. As vehicular emissions contribute different pollutants to the environment, due to the existence of different pollutants it will not be appropriate to see the belongingness of the level of pollution of a particular region to exactly one of the predefined clusters. Rather, it will be appropriate to see the membership values, full or partial, of the level of pollution of a particular region to one or more than one of the predefined clusters of pollution. So the application of a fuzzy technique would be convenient here. In the next paragraph, a literature review of some fuzzy applications has been put forward.

Zadeh (1965) developed the concept of fuzzy set theory (FST) particularly to deal with the situations pertaining to non-probabilistic uncertainty. A complete presentation of all aspects of FST is available in the work of Zimmermann (1991). The applications of FST in dealing with ambiguous problems where non-probabilistic uncertainty prevails have been reflected in the works of Dewit (1982) and Ostaszewski (1993). Park and Park (2010) developed a design to visualize the fuzzy set operations considering the

traditional Zadehian theory of fuzzy sets in which it was taken that there is no difference between the fuzzy membership function and the fuzzy membership value for the complement of a fuzzy set. Baruah (2011a, 2011b) has shown that the membership value of a fuzzy number can be expressed as the difference between the membership function and a reference function, and therefore the fuzzy membership function and the fuzzy membership value for the complement of a fuzzy set are not the same. Based on this concept, Das (2012) tried to modify the design of Park and Park (2010) and was able to overcome the limitations of their work by visualizing the complement of a fuzzy set in a correct manner.

Derring and Ostaszewski (1995) have explained in their research work a method of pattern recognition for risk and claim classification. Bezdek (1981) has discussed in his fuzzy c-means technique that the data to be analyzed must be in the form of numerical vectors called feature vectors, and the number of clusters must be predefined for obtaining the membership values of the feature vectors. Das (2013) tried the fuzzy c-means algorithm of Bezdek with three different distances namely Euclidean distance, Canberra distance and Hamming distance which revealed that out of the three distances, the algorithm produces the result fastest as well as the most expected when Euclidean distance is considered and the slowest as well as the least expected when Canberra distance is considered.

In the present work, we shall try to see the membership values, full or partial, of the level of pollution due to vehicular emissions of a particular city to one or more of some predefined clusters of pollution. For this purpose, we shall arbitrarily choose fifty (50) cities (Table - 3) and six (06) predefined clusters of pollution (Table - 2). Out of the different pollutants emitted by vehicles, we shall consider Carbon Monoxide (CO₂), Ozone (O₃) and Sulfur Dioxide (SO₂) which have the most adverse effects on the environment (Table - 1). Further, due to the existence of these three pollutants in vehicular pollution, the data used to represent it are in the form of numerical vectors. In our work, we shall apply the fuzzy c-means technique of Bezdek (1981) on a dataset representing vehicular pollution (Table - 3) to

obtain the membership values of pollution of a city to one or more of the six predefined clusters.

In Section – 2, we shall explain the fuzzy c-means algorithm of Bezdek. The application of this algorithm on vehicular pollution has been shown in Section - 3. In Section – 4, we shall show the findings and analysis of our work. Section - 5 consists of the conclusions.

2. Bezdek's fuzzy c-means algorithm

The fuzzy c-means algorithm due to Bezdek, which we have applied on a dataset of vehicular pollution has been explained through the following steps.

Step 1: Choose the number of clusters, c , $2 \leq c < n$, where n is the total number of feature vectors. Choose m , $1 \leq m < \alpha$. Define the vector norm $\| \cdot \|$, generally defined as the Euclidean distance,

$$\|x_k - v_i\| = \sqrt{\sum_{j=1}^p (x_{kj} - v_{ij})^2} \quad (1)$$

where x_{kj} is the j^{th} feature of the k^{th} feature vector, for $k = 1, 2, \dots, n$; $j = 1, 2, \dots, p$ and v_{ij} , j -dimensional centre of the i^{th} cluster for $i = 1, 2, \dots, c$; $j = 1, 2, \dots, p$; n , p and c being the total number of feature vectors, features in each feature vector and total number of clusters respectively.

Choose the initial fuzzy partition

$$U^{(0)} = [\mu_{s_i}^{(0)}(x_k)]_{1 \leq i \leq c, 1 \leq k \leq n}$$

Choose a parameter $\epsilon > 0$. This will help us to decide when exactly to stop the iteration. Set the iteration counting parameter l equal to 0.

Step 2: Compute the fuzzy cluster centers $\{v_i^{(l)}\}_{i=1,2,\dots,c}$ given by the formula

$$v_i^{(l)} = \frac{\sum_{k=1}^n (\mu_{s_i}^{(l)}(x_k))^m x_k}{\sum_{k=1}^n (\mu_{s_i}^{(l)}(x_k))^m}$$

for $i = 1, 2, \dots, c$; $k = 1, 2, \dots, n$.

(2)

Step 3: Compute the new partition matrix, i.e. membership matrix

$$U^{(l+1)} = [\mu_{s_i}^{(l+1)}(x_k)]_{1 \leq i \leq c, 1 \leq k \leq n}, \quad \text{where}$$

$$\mu_{s_i}^{(l+1)}(x_k) = \frac{1}{\sum_{j=1}^c \left(\frac{\|x_k - v_i^{(l)}\|}{\|x_k - v_j^{(l)}\|} \right)^{\frac{2}{m-1}}} \quad (3)$$

for $i = 1, 2, \dots, c$ and $k = 1, 2, \dots, n$. If $x_k = v_i^{(l)}$, formula (3) cannot be used. In this case the membership function is

$$\mu_{s_i}^{(l+1)}(x_k) = \begin{cases} 1 & \text{if } k=i \\ 0 & \text{if } k \neq i, i=1, 2, \dots, c \end{cases} \quad (4)$$

Step 4: Compute $\Delta = \|U^{(l+1)} - U^{(l)}\|$

If $\Delta > \epsilon$, repeat steps 2, 3 and 4. Otherwise, stop at some iteration count l^* . To make the result operational a fifth step had been introduced by Derring and Ostaszewski (1995).

Step 5: The final fuzzy matrix U^{l^*} is structured for operational use by means of the normalized α -cut, for some $0 < \alpha < 1$. All membership values less than α are replaced with zero and the function is renormalized to preserve partition condition.

3. Application on vehicular pollution

In our present work we have chosen 50 cities to see the level of pollution of a particular city due to vehicular emissions. Although vehicular emissions contribute different pollutants to the environment, we have considered only three, Carbon Monoxide (CO), Ozone (O₃) and Sulfur dioxide (SO₂), which have the most adverse effects on environment (Table - 1). Due to the existence of these three pollutants, unlike in the traditional techniques, it would not be appropriate to see the belongingness of the level of pollution of a particular city to exactly one of some predefined clusters. Rather, it would be appropriate to see the membership values, full or partial, of the level of pollution of a particular city to one or more of the predefined clusters of pollution. Therefore the application of a fuzzy technique will be more convenient here. For this purpose, we have divided the whole range of pollution into 6 clusters namely: Good (C1), Moderate (C2), Unhealthy for Sensitive Group (C3), Unhealthy (C4), Very Unhealthy (C5) and Hazardous (C6) (Table - 2).

Table 1: Vehicle Pollution Emission

Emission	Sources	Harmful Effects
Carbon Monoxide (CO)	Tailpipes	Human health, climate change
Ozone (O ₃)	NO _x and volatile organic hydrocarbons (VOC)	Human health, plants, aesthetics
Sulfur Dioxide (SO ₂)	Diesel vehicle tailpipes	Human health and ecological damage

Source: USEPA (2000), *Indicators of the Environmental Impacts of Transportation*, Center for Transportation and the Environment (www.itre.ncsu.edu/cte); ORNL, *Transportation Energy Data Book* ORNL (www.ornl.gov).

Table 2: Predefined clusters of pollutions

Cluster Name	Carbon monoxide (CO) (ppm)	Ozone (O ₃) (ppm)	Sulfur dioxide (SO ₂) (ppm)
Good	0 - 4	0 - 0.064	0 - 0.03
Moderate	> 4 - 9	0.065-0.084	> 0.03 - 0.14
Unhealthy for sensitive groups	> 9 - 12	0.085-0.104	> 0.14 - 0.22
Unhealthy	> 12 - 15	0.105-0.124	> 0.22 - 0.30
Very unhealthy	> 15 - 30	0.125-0.404	> 0.30 - 0.60
Hazardous	> 30 - 50	0.405-0.60	> 0.60 - 1.0

Source: Guidelines for the Reporting of daily Air Quality –
the Air Quality Index (AQI), U.S. Environmental Protection Agency (EPA)

In Table – 2, the ranges of the numerical values of each pollutant (can also be thought as feature) i.e. Carbon Monoxide (CO), Ozone (O₃) and Sulfur Dioxide (SO₂) have been provided for evaluations needed Bezdek's algorithm which we have applied on a dataset of vehicular pollution (Table - 3) for the purpose of partitioning it into 6 predefined clusters: Good(C1), Moderate (C2), Unhealthy for Sensitive Group (C3), Unhealthy (C4), Very Unhealthy (C5) and Hazardous (C6). Bezdek (1981) has discussed in his fuzzy c-means technique that the data to be analyzed must be in the form of numerical vectors called feature vectors, and the number of clusters must be predefined for obtaining the membership values of the feature vectors. In our present work, the data we have used to represent the pollution due to

vehicular emission of a city is in the form of numerical vector i.e. to explain it mathematically if X represents the pollution level of a city, then X is in the following form

$$X = (x_1, x_2, x_3),$$

where x_1 , x_2 and x_3 are the amounts (in ppm) of Carbon Monoxide (CO), Ozone (O₃) and Sulfur Dioxide (SO₂) respectively available in X (Table 3).

As the dataset we have analyzed satisfies the prerequisite conditions of Bezdek i.e. it is in the form of numerical vectors and also the number of clusters has been predefined, 6 in our case, the fuzzy c-means technique of Bezdek (1981) has been considered to be the most appropriate to be applied in the dataset of our present work.

Table 3: The dataset of pollution due to vehicular emission

City_ID	CO	O ₃	SO ₂	City_ID	CO	O ₃	SO ₂
1	4.000	0.060	0.020	26	1.000	0.060	0.010
2	4.000	0.070	0.050	27	2.000	0.061	0.011
3	10.000	0.086	0.190	28	3.000	0.062	0.012
4	13.000	0.109	0.250	29	4.000	0.063	0.013
5	17.000	0.126	0.400	30	15.000	0.124	0.300
6	35.000	0.408	0.650	31	16.000	0.125	0.310
7	36.000	0.409	0.651	32	17.000	0.126	0.320
8	18.000	0.127	0.401	33	18.000	0.127	0.330
9	14.000	0.110	0.251	34	19.000	0.128	0.340
10	16.000	0.087	0.191	35	20.000	0.129	0.350
11	5.000	0.071	0.051	36	21.000	0.130	0.360
12	4.000	0.061	0.021	37	13.000	0.230	0.250
13	5.000	0.062	0.022	38	14.000	0.231	0.251
14	6.000	0.072	0.520	39	15.000	0.232	0.252
15	12.000	0.088	0.192	40	16.000	0.233	0.253
16	15.000	0.111	0.252	41	25.000	0.234	0.254
17	19.000	0.128	0.402	42	3.000	0.065	0.030
18	50.000	0.600	1.000	43	4.000	0.066	0.031
19	5.000	0.067	0.040	44	5.000	0.067	0.032
20	6.000	0.068	0.050	45	6.000	0.068	0.033
21	7.000	0.069	0.060	46	7.000	0.069	0.034
22	8.000	0.070	0.070	47	8.000	0.070	0.035
23	9.000	0.071	0.080	48	49.000	0.590	0.950
24	10.000	0.072	0.150	49	3.000	0.067	0.253
25	11.000	0.073	0.151	50	15.000	0.234	0.250

4. Findings and analysis

In the traditional way of designating the level of pollution of a particular region or city, we have observed that out of the different available pollutants only one whose value is the maximum is considered, based on which the intensity of pollution of that particular city is calculated and accordingly placed to belong to exactly any one of some predefined clusters. In this method the existence or availability of other pollutants is completely neglected. In our present work, we

have observed that out of the three available pollutants i.e. CO, O₃ and SO₂, the availability of CO is the maximum for every city (Table - 3). Thus following the traditional way we have first considered the availability of CO out of the three available pollutants and have got the results as given in Table - 4. Here we have neglected the presence of the other two pollutants i.e. O₃ and SO₂, and therefore every city belongs to exactly any one of the 6 predefined clusters of pollution. A summarized form of the results in Table - 4 has

been given in Table - 5. But the fact is that we two pollutants.
cannot simply neglect the existence of the other

Table 4. Full membership values of the cities to exactly any one of the six(06) predefined clusters of pollution when the traditional way is followed.

FUZZY MEMBERSHIP VALUES							
CITY_ID	C1	C2	C3	C4	C5	C6	SUM
1	1	0	0	0	0	0	1
2	1	0	0	0	0	0	1
3	0	0	1	0	0	0	1
4	0	0	0	1	0	0	1
5	0	0	0	0	1	0	1
6	0	0	0	0	0	1	1
7	0	0	0	0	0	1	1
8	0	0	0	0	1	0	1
9	0	0	0	1	0	0	1
10	0	0	0	0	1	0	1
11	0	1	0	0	0	0	1
12	1	0	0	0	0	0	1
13	0	1	0	0	0	0	1
14	0	1	0	0	0	0	1
15	0	0	1	0	0	0	1
16	0	0	0	1	0	0	1
17	0	0	0	0	1	0	1
18	0	0	0	0	0	1	1
19	0	1	0	0	0	0	1
20	0	1	0	0	0	0	1
21	0	1	0	0	0	0	1
22	0	1	0	0	0	0	1
23	0	1	0	0	0	0	1
24	0	0	1	0	0	0	1
25	0	0	1	0	0	0	1
26	1	0	0	0	0	0	1
27	1	0	0	0	0	0	1
28	1	0	0	0	0	0	1
29	1	0	0	0	0	0	1
30	0	0	0	1	0	0	1
31	0	0	0	0	1	0	1
32	0	0	0	0	1	0	1
33	0	0	0	0	1	0	1
34	0	0	0	0	1	0	1
35	0	0	0	0	1	0	1
36	0	0	0	0	1	0	1
37	0	0	0	1	0	0	1
38	0	0	0	1	0	0	1
39	0	0	0	1	0	0	1
40	0	0	0	0	1	0	1
41	0	0	0	0	1	0	1
42	1	0	0	0	0	0	1
43	1	0	0	0	0	0	1
44	0	1	0	0	0	0	1
45	0	1	0	0	0	0	1
46	0	1	0	0	0	0	1
47	0	1	0	0	0	0	1
48	0	0	0	0	0	1	1
49	1	0	0	0	0	0	1
50	0	0	0	1	0	0	1
SUM	10	12	4	8	12	4	
Fuzzy C-means Clustering with 3 coverage data pattern after 0 iteration ,where epsilon =0.05							

Table 5. A city belonging to exactly one of the six predefined clusters.

Cluster Name	City_ID
Good (C1)	1, 2, 12, 26, 27, 28, 29, 42, 43, 49
Moderate (C2)	11, 13, 14, 19, 20, 21, 22, 23, 44, 45, 46, 47
Unhealthy for Sensitive Groups (C3)	3, 15, 24, 25
Unhealthy (C4)	4, 9, 16, 30, 37, 38, 39, 50
Very Unhealthy (C5)	5, 8, 10, 17, 31, 32, 33, 34, 35, 36, 40, 41
Hazardous	6, 7, 18, 48

Therefore we have applied the fuzzy c-means technique of Bezdek on our dataset (Table - 3) considering the existence of all three pollutants i.e. CO, O₃ and SO₂, and have tried to see the level of pollution of a city as a combined effect of all these three pollutants. The results obtained by this combined effect have been given in Table - 6. Here it can be observed that each city does not belong to exactly one of the six predefined clusters of pollution. Instead we find that some cities belong to exactly one while some cities belong to more than one of the six predefined clusters of pollution. Here, the membership values of the level of pollution of a city to a predefined cluster lies between 0 and 1, both 0 and 1 inclusive, unlike the traditional way where we find this value is either 0 or 1. Summarized results have been shown in Table - 7. In Figure – 1, we have provided a graphical representation of this result to show the membership value of the level of pollution of a city to a cluster of pollution by symbolizing the full membership value by a

diamond and partial membership value by a square.

If we observe the 3rd column of Table – 7, we find that the cities with partial membership values are in two consecutive clusters of pollution i.e. a city partially belong to two clusters which appear successively in the hierarchy of pollution. For example, the city with city_ID 11 belongs to cluster C1 (i.e. Good) with membership value 0.4817 and to cluster C2 (i.e. Moderate) with membership value 0.5183. This means that the pollution level of the city with city_ID 11 is neither fully ‘Good’ nor fully ‘Moderate’. It also indicates that if adequate measure is taken there still is approximately 48% possibility that this city could belong to cluster C1 (which is ‘Good’). Similar arguments can be given in support of the other cities appearing in column - 3 of Table - 7. The city_IDs and the names of the two different clusters where these belong partially have been shown in Table - 8. In Table – 8, we have provided the possibility of shifting the level of

pollution of a city to a cluster which has less average pollution out of the two clusters to which that city partially belongs. A graphical representation of the same has been given in Figure - 2.

Table 6. Membership values of the cities to one or more than one of the six predefined clusters of pollution when the availability of all the pollutants are considered.

AFTER TAKING α -CUT =0.2							
FUZZY MEMBERSHIP VALUES							
CITY_ID	C1	C2	C3	C4	C5	C6	SUM
1	1	0	0	0	0	0	1
2	1	0	0	0	0	0	1
3	0	0	1	0	0	0	1
4	0	0	0.6257	0.3743	0	0	1
5	0	0	0	1	0	0	1
6	0	0	0	0	0.4573	0.5427	1
7	0	0	0	0	0.408	0.592	1
8	0	0	0	0.5509	0.4491	0	1
9	0	0	0.2589	0.7411	0	0	1
10	0	0	0	1	0	0	1
11	0.4817	0.5183	0	0	0	0	1
12	1	0	0	0	0	0	1
13	0.4821	0.5179	0	0	0	0	1
14	0	1	0	0	0	0	1
15	0	0	1	0	0	0	1
16	0	0	0	1	0	0	1
17	0	0	0	0	1	0	1
18	0	0	0	0	0	1	1
19	0.4818	0.5182	0	0	0	0	1
20	0	1	0	0	0	0	1
21	0	1	0	0	0	0	1
22	0	1	0	0	0	0	1
23	0	0.4064	0.5936	0	0	0	1
24	0	0	1	0	0	0	1
25	0	0	1	0	0	0	1
26	1	0	0	0	0	0	1
27	1	0	0	0	0	0	1
28	1	0	0	0	0	0	1
29	1	0	0	0	0	0	1
30	0	0	0	1	0	0	1
31	0	0	0	1	0	0	1
32	0	0	0	1	0	0	1
33	0	0	0	0.5513	0.4487	0	1
34	0	0	0	0	1	0	1
35	0	0	0	0	1	0	1
36	0	0	0	0	1	0	1
37	0	0	0.6248	0.3752	0	0	1
38	0	0	0.2587	0.7413	0	0	1
39	0	0	0	1	0	0	1
40	0	0	0	1	0	0	1
41	0	0	0	0	1	0	1
42	1	0	0	0	0	0	1
43	1	0	0	0	0	0	1
44	0.4819	0.5181	0	0	0	0	1
45	0	1	0	0	0	0	1
46	0	1	0	0	0	0	1
47	0	1	0	0	0	0	1
48	0	0	0	0	0	1	1
49	1	0	0	0	0	0	1
50	0	0	0	1	0	0	1
SUM	11.9275	9.4789	6.3617	12.3341	6.7631	3.1347	
Fuzzy C-means Clustering with 3 coverage data pattern after 8 iteration ,where epsilon =0.05							

Table 7. A city belonging to one or more than one of the six predefined clusters.

Cluster Name	City_IDs which have full membership values	City_IDs with partial membership values
Good (C1)	1, 2, 12, 26, 27, 28, 29, 42, 43, 49	11 (.4817), 13 (.4821), 19 (.4818), 44 (.4819)
Moderate (C2)	14, 20, 21, 22, 45, 46, 47	11 (.5183), 13 (.5179), 19 (.5182), 23 (.4064), 44 (.5181)
Unhealthy for Sensitive Groups (C3)	3, 15, 24, 25	4 (.6257), 9 (.2589), 23 (.5936), 37 (.6248), 38 (.2587)
Unhealthy (C4)	5, 10, 16, 30, 31, 32, 39, 40, 50	4 (.3743), 8 (.5509), 9 (.7411), 33 (.5513), 37 (.3752), 38 (.7413)
Very Unhealthy (C5)	17, 34, 35, 36, 41	6 (.4573), 7 (.4080), 8 (.4491), 33 (.4487)
Hazardous	18, 48	6 (.5427), 7 (.592)

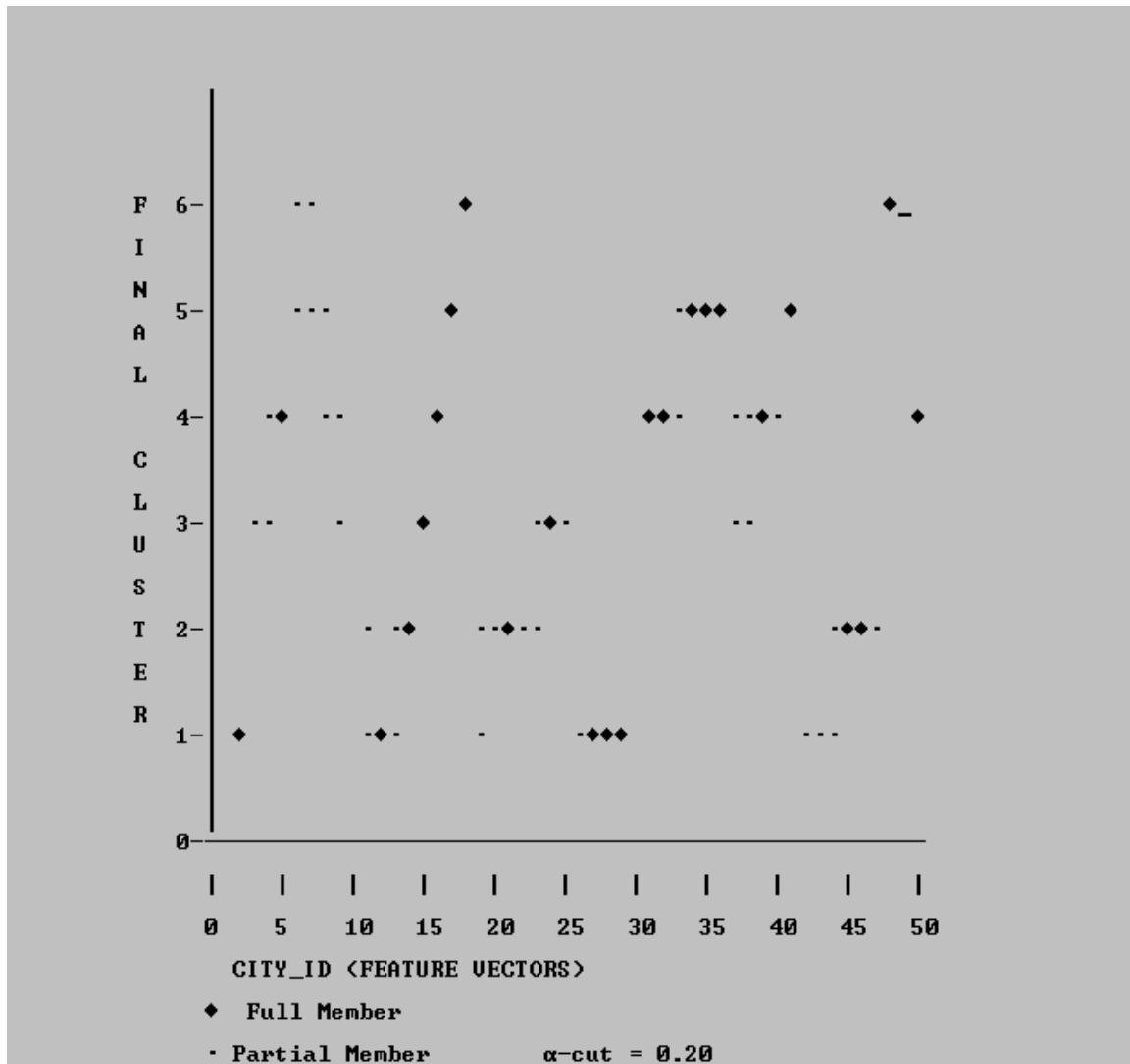


Figure 1. Partial and full membership values of the cities to one or more than one of the six (06) predefined clusters of pollution when the availability of all the pollutants are considered.

Table 8. City_IDs with respective possibility of shifting to a cluster with less average pollution

CITY_ID	CLUSTER NAMES WHERE THE CITIES ARE PARTIALLY PRESENT		POSSIBILITY OF SHIFTING TO A CLUSTER WITH LESS AVERAGE POLLUTION
	CLUSTER WITH LESS AVERAGE POLLUTION	CLUSTER WITH MORE AVERAGE POLLUTION	
4	Unhealthy for Sensitive Groups	Unhealthy	63%
6	Very Unhealthy	Hazardous	46%
7	Very Unhealthy	Hazardous	41%
8	Unhealthy	Very Unhealthy	55%
9	Unhealthy for Sensitive Groups	Unhealthy	26%
11	Good	Moderate	48%
13	Good	Moderate	48%
19	Good	Moderate	48%
23	Moderate	Unhealthy for Sensitive Groups	41%
33	Unhealthy	Very Unhealthy	55%
37	Unhealthy for Sensitive Groups	Unhealthy	62%
38	Unhealthy for Sensitive Groups	Unhealthy	26%

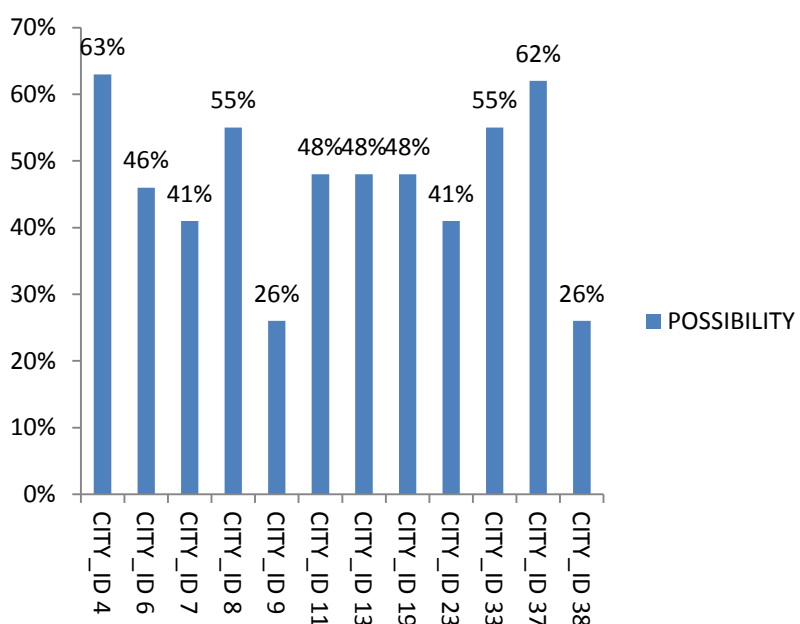


Figure2. City_IDs with respective possibility of shifting to a cluster with less average pollution

From Table - 8 and Figure – 2, we can see that it is possible to identify only those cities, for the purpose of improving its level of pollution, which belong partially to more than one clusters of pollution. As we have tried with the combined effect of all three pollutants, we are able to see some cities appearing partially in more than one clusters of pollution. In the traditional way it is not possible to identify such cities as each city belongs to exactly one out of the six predefined clusters.

5. Conclusions

Due to an almost exponential increase in the number of vehicles, vehicular pollution in the present scenario has become a major concern as it has the maximum impact on ground level population. In the traditional way of designating the level of pollution of a region or city due to vehicular emissions, out of the available pollutants only one whose availability is the maximum has been considered for calculations, based on which that particular city is placed to belong to exactly one of some predefined clusters of pollutions. In our work, we have applied the fuzzy c-means technique of Bezdek on a dataset of vehicular pollution with three pollutants- CO, O₃ and SO₂. Here unlike in the traditional way, we have considered the availability of all three pollutants, and have found that the level of pollutions of some cities belongs partially to more than one clusters of pollution. Due to the partial belongingness to more than one clusters of pollution, it is possible to identify those cities where there still is possibility of improving the level of pollution. Such identification is not possible when the traditional way is applied due to the lack of partial belongingness of cities to more than one clusters of pollution.

References

1. Zadeh, L. A. (1965), Fuzzy Sets, *Information and Control*, Vol.8, Issue 3, 338-353.
2. Zimmermann, H.J. (1991), Fuzzy Set Theory and its Applications, Second Edition, Kluwer Academic Publishers, Boston Massachusetts.
3. Dewit, G. W. (1982), Underwriting and Uncertainty, *Insurance: Mathematics and Economics*, Vol.1, Issue 4, 277-285.
4. Ostaszewski, K. (1993), An Investigation into Possible Applications of Fuzzy Sets Methods in Actuarial Science, Society of Actuaries, Schaumburg, Illinois.
5. Park, Y. and J. Park (2010), Interactive Visualization of Fuzzy Set Operations, *SPIE-IS&T*, Vol. 7530 753002, 1-12.
6. Baruah, H. K. (2011a), Towards Forming a Field of Fuzzy Sets, *International Journal of Energy, Information and Communications*, vol. 2, Issue 1, 16-20.
7. Baruah, H. K. (2011b), The Theory of Fuzzy Sets: Beliefs and Realities, *International Journal of Energy, Information and Communications*, vol. 2, Issue 2, 1-22.
8. Das, S. (2012), A New Approach to Visualization of Fuzzy Set Operations, *International Journal of Energy, Information and Communications*, Vol. 3, Issue 3, 29-40.
9. Derrig, R. A. and K. M. Ostaszewski (1995), Fuzzy Techniques of Pattern Recognition in Risk and Claim Classification, *Journal of Risk and Insurance*, Vol. 62, Issue 3, 447-482.
10. Bezdek, J.C. (1981), Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York.
11. Das, S. (2013), Pattern Recognition using the Fuzzy c-means Technique, *International Journal of Energy, Information and Communications*, vol. 4, Issue 1, 1-14.