Journal of Infrastructure, Policy and Development 2024, 8(5), 3926.
https://doi.org/10.24294/jipd.v8i5.3926

Article

Similarity analysis study of the carbon emission metrics through fuzzy entropy clustering

Sudha Prathyusha Jakkaladiki[1,*], Martina Janečková[2], Filip Malý[1]

[1] Department of Informatics and Quantitative Methods, Faculty of Informatics and Management, University of Hradec Kralove, 50003 Hradec Kralove, Czech Republic

[2] Department of Management, Faculty of Informatics and Management, University of Hradec Kralove, 50003 Hradec Kralove, Czech Republic

*** Corresponding author:** Sudha Prathyusha Jakkaladiki, sudha.jakkaladiki@uhk.cz

CITATION

Jakkaladiki SP, Janečková M, Malý F. (2024). Similarity analysis study of the carbon emission metrics through fuzzy entropy clustering. Journal of Infrastructure, Policy and Development. 8(5): 3926. https://doi.org/10.24294/jipd.v8i5.3926

ARTICLE INFO

Received: 3 January 2024
Accepted: 7 February 2024
Available online: 29 April 2024

Abstract: The world has changed to a massive degree in the past thousands of years. Most of the time, the amount of carbon dioxide in the atmosphere remains constant. In the late 18th century, according to the sources of CDIAC and NOOA, the level of carbon dioxide began to rise, and then in the 20th century, it went through the roof, reaching levels that had not been seen in nature for millions of years. The increase in carbon in the atmosphere is the major contributing factor to climate change. The key to reversing the damage is restoring the earth's delicate, balanced carbon cycle. As carbon cycle depicts the way carbon moves around the earth. It consists of sources that emit the carbon component into the atmosphere. The biological side of the carbon cycle is well balanced due to respiration, where carbon dioxide is released into the atmosphere, then plants, bacteria, and algae take carbon dioxide out of the atmosphere during photosynthesis and the process they use to generate chemical energy. On the other hand, oceans are the best sources and sinks; carbon dioxide is endlessly being absorbed into the ocean and released from the oceans almost exactly at the same rate, which is rapidly influencing the carbon cycle. Similarity is a methodology that has many applications in the real world. The current research article is destined to study how statistics of carbon emission metrics are alike and belong to one cluster. In the current study, the research is destined to derive a similarity analysis of several countries' carbon emission metrics that are alike and often fall in the range of [0, 1]. And deriving the proximity of the carbon emission metrics leading to similarity or dissimilarity. In the current context of data matrices of numerical data, an Euclidian measure of distance between two data elements will yield a degree of similarity. The current research article is destined to study the similarity analysis of carbon emission metrics through fuzzy entropy clustering.

Keywords: carbon emissions; similarity analysis; fuzzy entropy clustering; Euclidean distance; global warming; entropy of fuzzy sets

1. Introduction

Carbon emissions have been recognized as one of the main causes of global warming. The principal sources of carbon emissions are burning fossil fuels and forests, the production of cement, deforestation, and natural sources (respiration or decomposition of organisms, weathering of carbonate rocks, forest fires, etc.). For the present research, the focus is on carbon emissions caused by human activities. Since the industrial revolution more than 200 years ago, the concentration of carbon emissions has increased by 50% due to human activities (NASA, 2023). The reason behind its harmful effect is that it causes the greenhouse effect, therefore the earth is slowly warming (NASA, 2023).

The whole subject of further global warming prevention, energy efficiency, and sustainable resources has been getting more and more attention in the past years. Researchers emphasize the importance of investigation and overall discussion about these phenomena, which are influencing everyone on the planet (Chen and Wang, 2022; Gao et al., 2022, 2023). The number of research articles focusing solely on carbon emissions is also rapidly growing. In the Scopus database about 2000 articles with the keyword "carbon emissions" were published in 2003, while in 2023 it was almost 35000 articles. The authors focus on the evolution of CE in time, its price, impacts on the environment, causes, methods of measuring, etc., while emphasizing the importance of further research on this topic (Chen and Wang, 2022; Chen et al., 2022; Feng et al., 2018; Wang et al., 2023a; Wang et al., 2023b).

To decrease the production of carbon emissions, international organizations such as the United Nations and the European Union have taken precautions. In 2015 the Paris Agreement (COP21) was approved by the member states of the UN, where a long-term temperature goal was set and the 17 Sustainable Development Goals were established (United Nations General Assembly, 2015). Another precautionary measure was taken by the European Commission in 2019. The European Green Deal was approved to become the first climate-neutral continent in which the aim is to reduce net greenhouse gas emissions by at least 55% by 2030 (European Commission, 2023), Following on from previously concluded agreements, another conference—COP28 UN Climate Change took place in December 2023 in Dubai. An agreement was reached to accelerate emission reductions towards net zero by 2050. The deal includes an agreement to transition away from fossil fuels and to reduce global emissions by 43% by 2030 (European Commission, 2023).

Following the example of international institutions, individual organizations commit to goals by the reduction of the carbon footprint. Organizations such as Honeywell Forge, Orange, PwC, are committing to net zero, which means cutting greenhouse emissions to as close to zero as possible (Honeywell, 2023; Orange, 2023; PwC, 2023).

Despite all interventions and precautionary measures, the concentration of carbon dioxide is still rising (NASA, 2023). Achieving the set goals does not seem to be attainable to its full extent, which is described for example in the Emissions Gap Report 2022 (UN). Also as mentioned in the report the international community is falling far short of the Paris goals, with no credible pathway to 1.5 °C in place (United Nations, 2022). According to the European Environment Agency, the goals don't seem to be attainable to their full extent. The necessity to investigate challenges in carbon emissions forecasting and development is evident.

The current study is destined to determine the similarity analysis of carbon emission rates between different countries and the way that we can segregate those countries into different and distinct clusters to bring common policies that lead to common goals for the countries belonging to the same cluster.

2. Related work

In this section, important related research will be discussed from both content and methodological point of view. There are several different methods to measure carbon emissions production as well as to analyze and utilize the data collected by the methods mentioned above. Among the measurement methods are direct emission measurement, modeling, simulation, remote sensing technologies, continuous monitoring systems, etc., while these methods often complement each other. If we have a look at the methods of working with the data, it might be comparative analysis, cluster analysis, least squares method, remote sensing monitoring and analysis, etc. In one of the current research on this topic, comparative analysis was used to observe similarities between the level and the structure of the emissions released into the atmosphere (Brodny and Tutak, 2021). The purpose of the study was to group EU countries into homogenous groups by the amount of gas and dust emissions into the atmosphere. For measuring the evolution of carbon emissions production in time and space, integration of multisource remote sensing data was used while employing the partial least squares regression model (Wang et al., 2023a; Wang et al., 2023b). The results are highly useful in showing visually the evolution of CE production as well as the speed of the evolution in different areas. Although these methods bring many new findings, there is a lack of research on carbon emissions production with the use of the similarity analysis method, which may shed light on similarities and dissimilarities in CE emissions among different countries of the world.

In similarity analysis, the data can be crisp or fuzzy. Since the nature of data is fuzzy (not only integers) focusing on the current area of research of similarity analysis, but also as fuzzy entropy clustering (which is one of the similarity analysis methods). Since the purpose of fuzzy clustering is to group data points into multiple distinct clusters, it is useful in domains such as medical image analysis, pattern recognition and image processing, uncertainty and inaccuracy analysis, environmental modeling, bioinformatics, stock price prediction, sales forecasting, etc. (Chang et al., 2009; Enke and Mehdiyev, 2013; Jakkaladiki and Janečková, 2023; Vatansever et al., 2020). In the research article the degree of uncertainty of the data has been calibrated through fuzzy entropy methodology and the degree of uncertainty of the data set has been derived. In the current article, the fuzzy entropy clustering method will be discussed for CE production measuring and comparison for several countries.

Although there is reasonable research has been conducted on similarity analysis most of the methodologies are aimed at crisp data sets hence the current research article is intended to illuminate light on the similarity analysis on fuzzy data sets which is essentially derived from entropy-based fuzzy clustering.

3. Proposed methodology

A potential application of fuzzy sets is fuzzy clustering. Clustering is a powerful tool for data mining, and the purpose of data mining is to extract useful information from a data set. The mechanism of clustering takes place based on the concept of similarity. Mathematically, similarity means two similar points (data elements) belong to the same cluster, and vice versa, two dissimilar data elements belong to two different clusters. The clustering can be either crisp or fuzzy; in crisp data sets, there are well-defined boundaries, whereas, in fuzzy data sets, the boundaries are vague.

The algorithm of entropy-based fuzzy clustering depicts (**Figure 1**) that entropy is an index determined based on the similarity measure, which is based on the Euclidean distance. Post applying the principle, the data element with the minimum entropy value is selected as the cluster center, and the data points with a similarity value with the cluster center greater than the specified value will form a cluster. The process involves the data elements in a multi-dimensional form, and the objective is to identify the cluster center, which determines the entropy that measures similarity based on the numerical value of Euclidian distance. For example, there is an element *z*, and from *j*, the Euclidean distance is D_{zi}, and the similarity between two data elements is S_{ji}. The distance between two data elements is inversely proportional to the similarity of the data elements; in a way, the derivation of the Euclidian distance will lead to the path of similarity between those entities.

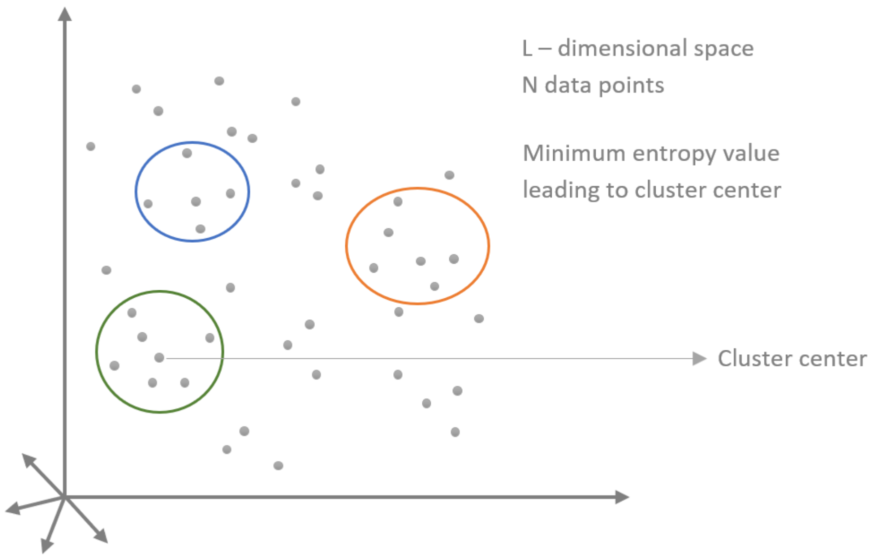

Figure 1. Entropy-based fuzzy clustering (author's elaboration).

N number of data points and $N = 1000$ and L-dimensional space and to represent *L* number numerical values. Identification of cluster center follows the below algorithm. For N data elements in L-D hyperspace.

Step 1: Sorting the data set in *N* rows and *L* columns. *N* number of data elements and each data element has *L* dimension leads to matrix L × N.

Step 2: Determine the Euclidian distance between the data elements *I* and *J*. Euclidean distance is mentioned below $D_{ij} = \sqrt{\sum_{k=1}^{L}}(x_{ik} - x_{jk})\,2$ where *K* varies from 1 to *L*.

Step 3: Determine the similarity S_{ij} between data points *I* and *J*.

$$s_{ij} = e^{-\propto d_{ij}} = 1/e^{\infty d_{ij}}$$

- If d_{ij} is more than 1/e to the power of alpha d_{zi}.
- If d_{ij} is higher than the value, then the similarity will be lower. Where α is constant to be determined.

To determine the value of α assume a similarity of 0.5 when the distance between two data points d_{ij} becomes equal, to the mean distance of all pairs of data elements.

$$d_{ij} = \bar{d} = 1/nc2 \sum_{i=1}^{n} \sum_{j>1}^{n} d_{ij} \tag{1}$$

(Source: Pratihar, 2014).

Therefore, get $\alpha = \ln 2/\bar{d}$, Note: S_{ij} varies in the range of 0.0 to 0.1.

If we have 5 data elements.

- $D_{21} = D_{12}$ $D_{41} = D_{14}$ (as distance is equal)
- Diagonal elements distance $D_{11} = D_{22} = D_{33} = D_{44} = D_{55} = 0$,

On further analysis of roadmap D_{12} D_{13} D_{14} D_{15} D_{23} D_{24} D_{25} D_{34} D_{35} D_{45} distance traveled is the sum of all 10 values for 5 data elements (**Figure 2**).

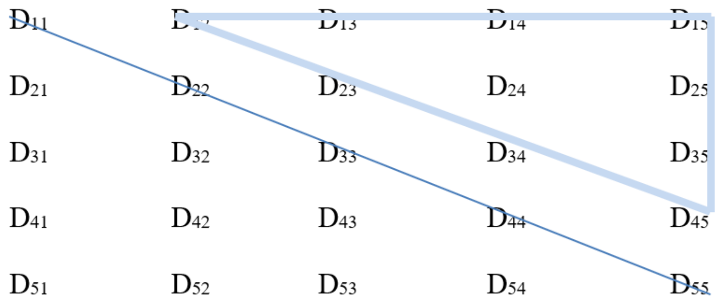

Figure 2. Representation of distance of data elements.

The process is to find the mean distance: $\alpha = \log 2/\bar{d}$, $s_{ij} = e^{-\propto d_{ij}} = 1/e^{\infty d_{ij}}$, $½ = e^{-\propto \bar{d}}$ by Considering Log for both sides. $\frac{\log 1}{2} = e^{-\bar{d}}$, $\log 1 - \log 2 = -\alpha\bar{d} \times \log e$ ($\log e = 1$) (Where $\bar{d}$ is mean distance) Since log 1 is 0 $-\log 2 = -\alpha\bar{d}$, $\alpha = \log 2/\bar{D}$, derivation of a constant value that is $\alpha = \log 2/\bar{D}$, leading to the α value that determines similarity and distance.

Step 4: Determining the entropy E of all N data elements.

$$E = -s\log s - (1 - s)\log(1 - s) \tag{2}$$

where E is the entropy of given data elements and S is the similarity of the two data elements d_{ij} for $S = 0$ means similarity between d_{ij} means the distance between two elements is S very high (Too Far) entropy becomes

$$E = -0\log s - (1 - 0)\log(1 - 0) = 0$$

Entropy is 0, if s = 1 $E = -1\log 1 - (1 - 1)\log(1 - 1) = 0$ again Entropy is 0 if the distance between $d_{ij} = 1$ For the two extreme values $S = 1$ Entropy became 0, $S = 0$ Entropy became 0 (**Figure 3**).

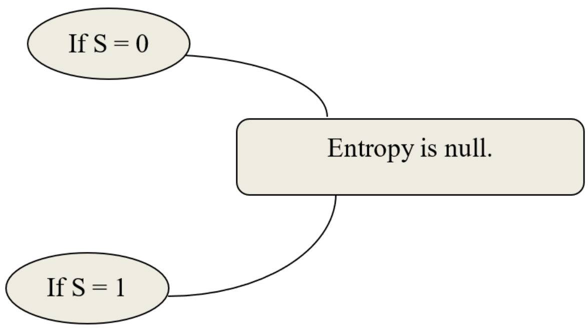

Figure 3. Relationship of similarity gradient and entropy (author's elaboration).

Post substituting S = 0.5 = $-½\log½ - ½\log½ = +1$ Entropy became 1, Based on the above philosophy the relationship is derived.

Step 5: The total entropy of each of the data elements X concerning all other data elements.

$$E_i = -\sum_{j \in x}^{j \neq i} (s_{ij} \log s_{ij} + (1 - s_{ij}) \log(1 - s_{ij})) \quad (3)$$

(Source: Pratihar, 2014).

where E_i = Entropy of the *i*-th Data element, once we attain the entropy of the elements then we can derive a clustering algorithm.

The following are the activities involved in determining the clustering based on the entropy:

Step 1: Determining Entropy E_i for each Data element x_i residing in [T] Hyperspace. N number of data elements in L-dimensional space.

Step 2: Derive x_i that has minimum E value and select it (x_i, min) as the cluster center.

Step 3: Substitute X_i min and the data points having similarity with x_i, min greater than β (Threshold value of similarity) in a cluster and remove them from [T].

Step 4: Check if [T] is empty if yes terminate the iteration, else navigate to Step 2 Concept of outliers: -Post clustering is completed, calibration of count the number of data elements. Residing in each cluster and if this number becomes greater than or equal to γ% of total data elements, declaring that this cluster is valid. 10% of 1000 is 100. When there is need to have 100 points in the cluster. The number of outliers is as minimal as possible ideal condition outliers will be 0.

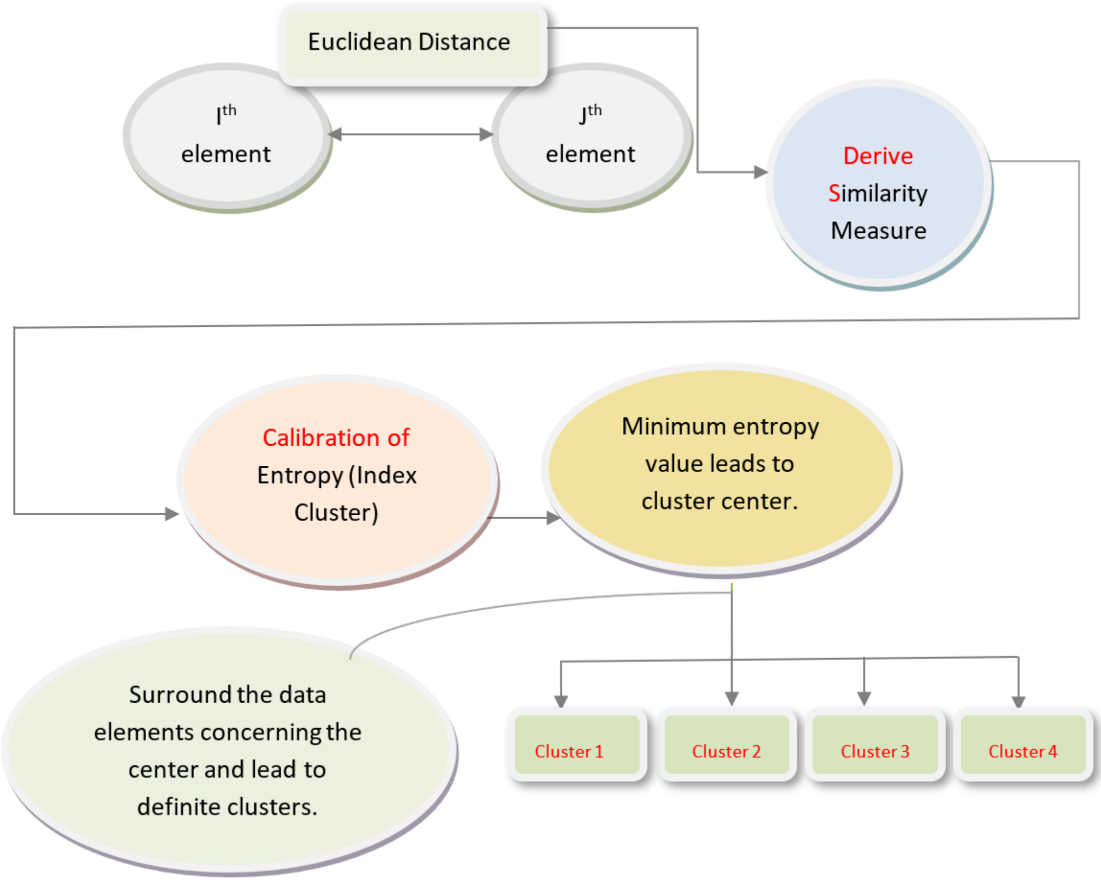

Figure 4. Ecosystem of similarity analysis through entropy-based clustering (author's elaboration).

Post clustering, the process will lead to deriving the count of data points residing in each cluster, and if this number becomes greater than or equal to γ% of the total number of data points, leading to a valid cluster otherwise these data points will be declared as an outlier. The whole process is depicted in **Figure 4**.

4. Experiment

Fuzzy clustering is one of the potential applications of fuzzy sets. Fuzzy clustering is a powerful tool for data analytics and also for mining and clustering involves grouping data elements into several clusters, which has been carried out based on the similarity of the data elements. There are several methods of fuzzy clustering, Fuzzy C means clustering algorithm, potential-based clustering, entropy-based clustering, and others. The current research article is being carried out based on the principle of fuzzy entropy clustering to derive distinct clusters and fuzzy clustering is being carried out with the concept of similarity and two similar points belonging to the same cluster.

The design of clustering is calibrated based on the relative distance of the data points with the cluster center and measuring Euclidian distance can be crucial as the distance of the data elements is a significant factor that influences the degree of similarity.

The following process has been followed to find the Euclidean distance which in turn derives the similarity of the data elements and finds the value of entropy of the data set (**Table 1**).

Table 1. List of countries.

Sq. No.	Country	2021 (mt)
P0	Saint Helena, Ascension and Tristan da Cunha	0.02
P1	Falkland Islands	0.05
P3	Kiribati	0.08
P4	Saint Pierre and Miquelon	0.09
P5	Cook Islands	0.11
P6	São Tomé and Príncipe	0.15
P7	Tonga	0.17
P8	British Virgin Islands	0.18
P9	Dominica	0.19
P10	Vanuatu	0.22

Annotation: mt—metric tons (Source: Kaggle, 2023)

Carry out fuzzy clustering based on the similarity and entropy values assume threshold value, i.e., beta = 0.5, to determine any such out Lier nothing but γ = 10%.

Determination of the Euclidian distance here we have data elements I e N = 10.

$$D_{ij} = \sqrt{\sum_{k=1}^{L}} (x_{ik} - x_{jk})\,2 \tag{4}$$

Calculating the Euclidian distance between two data elements *I*, *j* as several data elements is 10. There will be 10 c2 = 45 distance values (d_{ij}), As we know $d_{ji} = d_{ij}$ and diagonal elements d_{00}, d_{11}, d_{22}, d_{33}, d_{44}, d_{55} = 0 (diagonal elements are 0).

Mean of distance values = $\overline{d} = \sum d_{ij}/45$

Euclidian distance:

$$d(i,j) = \sqrt{\left(x_{i1} - x_{j1}\right)^2 + (x_{i1} - x_{j1})^2 + \cdots + (x_{in} - x_{jn})^2} \quad (5)$$

(Source: Pratihar, 2014).

D_{ij} = D0 to 9 = (0.05 − 0.02) = 0.03, 0.08 − 0.02 = 0.06, 0.09 − 0.02 = 0.07, 0.11 − 0.02 = 0.09 = 0.15 − 0.02 = 0.13, 0.17 − 0.02 = 0.15, 0.18 − 0.02 = 0.16,

Mean distance values $\bar{D} = \bar{D} = \sum d_{ij}/45 = 7.16/45 = 0.1591$

Alpha = ln2 /d bar = ln2/0.1591 = 0.69314/0.1591 = 4.35

Similarity between the points i and j

$$s_{ij} = e^{-\propto d_{ij}} = 1/e^{\infty d_{ij}}$$

$$S_{ij} = 1/2.718\ (4.35 \times 0.06)$$

These are the Euclidean distance, similarity gradient, and entropy values are calibrated based on the algorithm mentioned (**Table 2**).

Table 2. Calculations of Euclidean distance, similarity, and entropy for the selected countries.

D0	**ED**	**SG**	**E**	**D1**	**ED**	**SG**	**E**
0, 1	0.03	0.87766	0.371561	1, 0	0.03	0.87766	0.371561
0, 2	0.06	0.7703	0.538914	1, 2	0.03	0.87766	0.371561
0, 3	0.07	0.73751	0.575645	1, 3	0.04	0.84031	0.43915
0, 4	0.09	0.67606	0.629871	1, 4	0.06	0.7703	0.53891
0, 5	0.13	0.5681	0.683843	1, 5	0.1	0.64729	0.64918
0, 6	0.15	0.52077	0.692284	1, 6	0.12	0.59336	0.67561
0, 7	0.16	0.49861	0.693143	1, 7	0.13	0.5681	0.68384
0, 8	0.17	0.47738	0.692124	1, 8	0.14	0.54392	0.68928
0, 9	0.2	0.41898	0.679961	1, 9	0.17	0.47738	0.69212
D2	**ED**	**SG**	**E**	**D3**	**ED**	**SG**	**E**
2, 0	0.06	0.7703	0.538914	3, 0	0.07	0.73751	0.575645
2, 1	0.03	0.87766	0.371561	3, 1	0.04	0.84031	0.43915
2, 3	0.01	0.95743	0.17602	3, 2	0.01	0.95743	0.17602
2, 4	0.03	0.87766	0.575645	3, 4	0.02	0.91668	0.286804
2, 5	0.07	0.73751	0.575645	3, 5	0.06	0.7703	0.538914
2, 6	0.09	0.67606	0.629804	3, 6	0.08	0.70612	0.6055
2, 7	0.1	0.64729	0.64918	3, 7	0.09	0.67606	0.629871
2, 8	0.11	0.61974	0.66419	3, 8	0.1	0.64729	0.66419
2, 9	0.14	0.54392	0.68928	3, 9	0.13	0.5681	0.68384
D4	**ED**	**SG**	**E**	**D5**	**ED**	**SG**	**E**
4, 0	0.09	0.67606	0.629804	5, 0	0.13	0.5681	0.68384
4, 1	0.06	0.7703	0.538914	5, 1	0.1	0.64729	0.64918
4, 2	0.03	0.87766	0.371561	5, 2	0.07	0.73751	0.57564
4, 3	0.02	0.91668	0.286804	5, 3	0.06	0.7703	0.53891
4, 5	0.04	0.84031	0.43915	5, 4	0.04	0.84031	0.43915
4, 6	0.06	0.7703	0.538914	5, 6	0.02	0.91668	0.286804
4, 7	0.07	0.73751	0.575645	5, 7	0.03	0.87766	0.371561

Table 2. (*Continued*).

D4	ED	SG	E	D5	ED	SG	E
4, 8	0.08	0.70612	0.6055	5, 8	0.04	0.84031	0.43915
4, 9	0.11	0.61974	0.66419	5, 9	0.07	0.79419	0.575645
D7	**ED**	**SG**	**E**	**D8**	**ED**	**SG**	**E**
7, 0	0.16	0.49861	0.69314	8, 0	0.17	0.47738	0.69212
7, 1	0.13	0.5681	0.68384	8, 1	0.14	0.54392	0.68928
7, 2	0.1	0.64729	0.64918	8, 2	0.11	0.61974	0.66419
7, 3	0.09	0.67606	0.629871	8, 3	0.1	0.64729	0.66419
7, 4	0.07	0.79419	0.575645	8, 4	0.08	0.70612	0.6055
7, 5	0.03	0.87766	0.371561	8, 5	0.04	0.84031	0.43915
7, 6	0.01	0.95743	0.17602	8, 6	0.02	0.91668	0.286804
7, 8	0.01	0.95743	0.17602	8, 7	0.01	0.95743	0.17602
7, 9	0.04	0.84031	0.43915	8, 9	0.03	0.87766	0.371561
D6	**ED**	**SG**	**E**	**D9**	**ED**	**SG**	**E**
6, 0	0.15	0.52077	0.69228	9, 0	0.2	0.41898	0.67996
6, 1	0.12	0.59336	0.67561	9, 1	0.17	0.47738	0.69212
6, 2	0.09	0.67606	0.629871	9, 2	0.14	0.54392	0.68928
6, 3	0.08	0.70612	0.6055	9, 3	0.13	0.5681	0.68384
6, 4	0.06	0.7703	0.53891	9, 4	0.11	0.61974	0.66419
6, 5	0.06	0.7703	0.53891	9, 5	0.07	0.79419	0.575645
6, 7	0.01	0.95743	0.17602	9, 6	0.05	0.80454	0.49404
6, 8	0.02	0.91668	0.286804	9, 7	0.04	0.84031	0.43915
6, 9	0.05	0.80454	0.494	9, 8	0.03	0.87766	0.371561

The term Entropy is index and that in turn decides the cluster center.

Entropy of the *I*-th data element = $\sum_{j\in x}^{j\neq i}(s_{ij}\log s_{ij} + \left(1 - s_{ij}\right)\log(1 - s_{ij}))$

Entropy values are derived to be as follows:

Post calibration the cluster is determined based on the cluster center and below are metrics that have been calculated (**Tables 3** and **4**, **Figure 5**).

Table 3. Entropy measures.

Data	Entropy
E0	5.557346
E1	5.111212
E2	4.870239
E3	4.599934
E4	4.650482
E5	4.55988
E6	4.637905
E7	4.394427
E8	4.588815
E9	5.289786

Table 4. Clustering.

D7	Euclidian Distance	Similarity Gradient	Entropy	Cluster
7, 0	0.16	0.49861	0.69314	C1
7, 1	0.13	0.5681	0.68384	C1
7, 2	0.1	0.64729	0.64918	C1
7, 3	0.09	0.67606	0.629871	C1
7, 4	0.07	0.79419	0.575645	C1
7, 5	0.03	0.87766	0.371561	C2
7, 6	0.01	0.95743	0.17602	C2
7, 8	0.01	0.95743	0.17602	C2
7, 9	0.04	0.84031	0.43915	C2

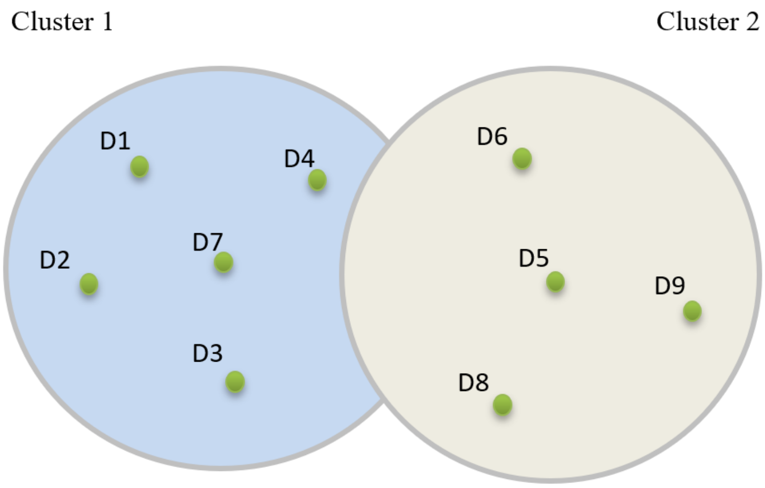

Figure 5. Cluster 1 and cluster 2: the outcome of the similarity analysis (author's elaboration).

5. Results and discussion

The purpose of the analysis of similarity is to find the statistical difference between the data elements and on the other hand. In the stream of data analytics, one of the most common reasons for cluster analysis is to derive members of each cluster technically called "data segmentation", in centroid clustering, the number of clusters is chosen and the determination of centroid for each cluster is based on the Euclidean distance and then each cluster is being determined based on the Entropy of the data elements. Based on the lab results, post applying the algorithm the cluster center has been chosen as D7 for cluster 1 and D5 of cluster 2, and the remaining data elements have been surrounded based on the entropy of the data set and several iterations to determine the optimum groupings. The algorithm will be iterated till the optimum results have been attained.

In the atmosphere Increasing carbon oxide leads to increases in average temperature around the world. Most of the carbon emissions come from burning fossil fuels come from carbon underground. Anything powered by fossil fuel releases carbon dioxide into the atmosphere. However, greater energy conservation and energy efficiency lead to lower levels of carbon dioxide in the air. There is a greater need to determine the number of gases such as carbon dioxide and there is a direct impact on using energy for traveling and powering homes and there is a direct impact on the energy used to create and in the developed world transport is the big components of the carbon footprint. The more sustainable energies like solar energy and wind power. recycling can also be a regular practice to reduce the carbon footprint, for the carbon neutral societies. The importance of clustering in the carbon footprint is to bring common policies to the countries belonging to the same cluster and to make an impact on the global level and definite course of methods and actions are selected by the government and institutions and groups of individuals from among alternatives and in the light of given conditions to guide and usually to determine present and future decisions. To apply the plan of action to work on the carbon footprint reduction.

6. Conclusion

Earth is the only planet that contains life. Based on the carbon emission metrics. The human impact on the planet is very dangerous and human-induced climatic change is the largest most pervasive threat to the natural environment and society. and climate change is getting abnormal. Human activities are inflating the carbon life cycle in the atmosphere at record rates. Restricting the radiation of the heat into the infinite space contributes to the overall warming of the planet. Invisible as they are carbon emissions can be hard to wrap heads around. Geo-dropping metrics demonstrate precise mention of where most of the critical carbon emissions are released and how metrics rapidly change over a single year. The available statistics provide insights on the consequences of carbon emission rates and there should be a change in policies to combat climate change.

The entropy-based clustering algorithm has its significant existence in determining the classification of cluster performance for multi-dimensional data sets and accelerates computational efficiency, effectuating in world real-world high-dimensional sets.

The number and nature of clusters depend upon the threshold value of similarity and are more flexible when compared to the fuzzy C means algorithm based on the research carried out entropy-based clustering yields more distinct but less compact clusters and the performance of clustering algorithms is data-dependent. In the current research, the fuzzy entropy clustering depends upon several factors such as α determining the relationship between Euclidian distance and the similarity and β threshold value of similarity and then γ denotes the number of outliers. Further research is being carried out based on the integration of the merits of the fuzzy C Means algorithm and the merits of the fuzzy entropy-based algorithm so that optimum performance can be achieved by eliminating inherent demerits.

Author contributions: Conceptualization, SPJ; state of art, MJ; methodology, SPJ; formal analysis, SPJ; theoretical formalism, MJ; investigation; SPJ and MJ; resources, references, MJ; computations, SPJ; data curation, MJ and SPJ; original draft preparation, SPJ and MJ; writing, review and editing, SPJ and MJ; Visualization, MJ and SPJ; supervision, FM. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: The intention of the article was expressed and proposed within the framework of doctoral studies, the study is supported by the SPEV project run by the faculty of informatics and management at the University of Hradec Kralove, Czech Republic.

Conflict of interest: The authors declare no conflict of interest.

References

Brodny, J., & Tutak, M. (2021). The analysis of similarities between the European Union countries in terms of the level and structure of the emissions of selected gases and air pollutants into the atmosphere. Journal of Cleaner Production, 279, 123641. https://doi.org/10.1016/j.jclepro.2020.123641

Chang, P. C., Liu, C. H., & Fan, C. Y. (2009). Data clustering and fuzzy neural network for sales forecasting: A case study in printed circuit board industry. Knowledge-Based Systems, 22(5), 344–355. https://doi.org/10.1016/j.knosys.2009.02.005

Chen, L., & Wang, K. (2022). The spatial spillover effect of low-carbon city pilot scheme on green efficiency in China’s cities: Evidence from a quasi-natural experiment. Energy Economics, 110, 106018. https://doi.org/10.1016/j.eneco.2022.106018

Chen, P., Vivian, A., & Ye, C. (2022). Forecasting carbon futures price: A hybrid method incorporating fuzzy entropy and extreme learning machine. Annals of Operations Research, 313(1), 559–601. https://doi.org/10.1007/s10479-021-04406-4

Enke, D., & Mehdiyev, N. (2013). Stock Market Prediction Using a Combination of Stepwise Regression Analysis, Differential Evolution-based Fuzzy Clustering, and a Fuzzy Inference Neural Network. Intelligent Automation & Soft Computing, 19(4), 636–648. https://doi.org/10.1080/10798587.2013.839287

Feng, Z., Tang, W., Niu, Z., & Wu, Q. (2018). Bi-level allocation of carbon emission permits based on clustering analysis and weighted voting: A case study in China. Applied Energy, 228, 1122–1135. https://doi.org/10.1016/j.apenergy.2018.07.005

Gao, D., Li, G., & Li, Y. (2022). Government cooperation, market integration, and energy efficiency in urban agglomerations—Based on the quasi-natural experiment of the Yangtze River Delta Urban Economic Coordination Committee. Energy & Environment, 33(8), 1679–1694. https://doi.org/10.1177/0958305X211047480

Gao, D., Zhou, X., Yan, Z., & Mo, X. (2023). The Impact of the Resource-Exhausted City Program on Manufacturing Enterprises’ Performance: A Policy Measures Perspective. Systems, 11(8), 415. https://doi.org/10.3390/systems11080415

Jakkaladiki, S. P., & Janečková, M. (2023). Study of Uncertainty and Inaccuracies of the Income for Inequality Paradox through Fuzzy Logic for Sustainable Development. Procedia Computer Science, 225, 564–573. Available online: https://doi.org/10.1016/j.procs.2023.10.041 (accessed on 3 November 2023).

Kaggle. (2023). Available online: https://www.kaggle.com/datasets/mabdullahsajid/tracking-global-co2-emissions-1990-2023 (accessed on 3 November 2023).

Pratihar, D. K. (2014). Soft computing: Fundamentals and applications (Neuaufl.). Alpha Science International Ltd.

Vatansever, M., Demir, İ., & Hepşen, A. (2020). Cluster and forecasting analysis of the residential market in Turkey: An autoregressive model-based fuzzy clustering approach. International Journal of Housing Markets and Analysis, 13(4), 583–600. https://doi.org/10.1108/IJHMA-11-2019-0110

Wang, M., Wang, Y., Teng, F., & Ji, Y. (2023a). The spatiotemporal evolution and impact mechanism of energy consumption carbon emissions in China from 2010 to 2020 by integrating multisource remote sensing data. Journal of Environmental Management, 346, 119054. https://doi.org/10.1016/j.jenvman.2023.119054

Wang, Y., Wang, M., Teng, F., & Ji, Y. (2023b). Remote Sensing Monitoring and Analysis of Spatiotemporal Changes in China’s Anthropogenic Carbon Emissions Based on XCO2 Data. Remote Sensing, 15(12), 3207. https://doi.org/10.3390/rs15123207