ISSN 2522-3488 Vol. 7, Issue 9: 42-47, DOI: 10.25236/IJNDES.2023.070908

Adaptive Region Classification and Prediction of Underwater Navigation Based on Data-driven

Xiangyu Sun

School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan, 114051, China

Abstract: *With the continuous improvement of sensor technology, machine learning and data processing capabilities, underwater navigation gradually turns to the data-driven model. In this paper, we make full use of the data in the appendix to make a reasonable judgment on the classification and prediction of underwater adaptive areas, and evaluate the performance. K-means clustering method is used to cluster the characteristics of gravity anomalies in different regions. When the number of clustering centers is set from 2 to 11:00, the clustering results show that when the number of clusters is 2, 3, 5, 7, 8, 9, all 100 iterations converge, and the clustering effect is good. When the number of clustering is 2, 3 and 5, the sum of the best distance is larger, and the distribution of contour coefficient has negative expansion; when the number of clustering is 7 and 8, the distribution of contour coefficient has negative expansion; when the number of clustering is 9, the contour coefficient is larger, the distribution is uniform on the front, the part of the contour coefficient less than 0 is very small, the convergence in the iterative process, the optimal distance and the sum of the optimal distance is relatively small: 63711.2. Therefore, when the number of clusters is 9, the fitting effect of each region is the best. Mark subclass 1 as area 1, mark subclass 2 as area 2, and so on.*

Keywords: *K-means clustering, Data-driven, Underwater Navigation, Adaptive region Classification*

1. Introduction

Underwater navigation is a key technology, which is widely used in the fields of ocean engineering, underwater archaeology, Marine resource exploration and Marine scientific research. In the past few decades, with the continuous improvement of sensor technology, machine learning and data processing capabilities, underwater navigation has gradually shifted to a data-driven model. This model provides new possibilities for adaptive area classification and prediction in underwater navigation by utilizing large-scale underwater data acquired by sensors, combined with advanced algorithms and models[1-3].

Traditional underwater navigation technologies, which usually rely on sensors such as prior maps, inertial navigation and sonar, have encountered some challenges in the underwater environment, such as inadequate adaptability to changes in the Marine environment, limited positioning accuracy and the need for real-time feedback. However, the data-driven approach, based on data acquired from a variety of sensors, utilizes machine learning, deep learning and data mining techniques to better address these challenges[4].

Adaptive area classification and prediction in underwater navigation is an important task[5-7]. Adaptation zones usually refer to specific areas in the underwater environment, which may include shallow seas, deep seas, coral reefs, rocky terrain, etc. Classifying and predicting these zones can help improve the adaptability and efficiency of underwater vehicles, submersibles, or underwater equipment in different environments. In addition, accurate classification and prediction of the adaptive areas also provide important support for Marine resource exploration, ecological protection and Marine scientific research.

Data-driven adaptation area classification and prediction rely on the collection, collation and analysis of large-scale underwater data. This includes data from sonar, lidar, cameras and other sensors, covering underwater terrain, water quality, biological information and more. From this data, machine learning models and deep neural networks can be used to identify, classify and predict different areas of adaptation. At the same time, the data-driven approach can be modified and improved with real-time feedback, thus improving the accuracy and reliability of the classification and prediction of the fit zones.

The research on the classification and prediction of underwater navigation adaptive area based on

ISSN 2522-3488 Vol. 7, Issue 9: 42-47, DOI: 10.25236/IJNDES.2023.070908

data drive provides a more accurate and efficient tool for the successful execution of underwater missions. However, there are still some challenges in this field, such as data quality, algorithm robustness and model generalization ability. Therefore, future research needs to continuously optimize algorithms, improve data quality, and continuously validate and refine models to better achieve adaptive area classification and prediction in underwater navigation.

Based on the background analysis, this paper intends to solve the following problems:

According to this, a set of gravity anomaly reference data A with a resolution of 1 / 1 (the distance between adjacent points is 1') is given.

Refine the reference map as much as possible, divide the region reasonably, and complete the adaptation and calibration of each region.

2. Research Method

According to the background of the subject, due to the different distribution of gravity anomaly characteristics in different regions, on the basis of gravity anomaly data, the K-means clustering method is used to cluster the gravity anomaly characteristics in different regions, and the similar gravity data are classified into one class. The area where the gravity data is located is classified into one category, and the label of the region is set. First of all, the number of multi-clusters is set, and the distribution of the contour coefficients of each cluster center, the sum of the best distance and whether it converges in the iterative process are calculated. Then compare the three, select the optimal cluster number of the screening target, which is the distribution of the contour coefficient is more uniform, the contour coefficient is greater than 0.8 and less than the negative value, the sum of the best distance and iterative convergence, and the optimal cluster number has the highest adaptability. Finally, mark subclass 1 as area 1, mark subclass 2 as area 2, label 2, and so on.

3. The establishment and solution of the model

According to the background of the topic, due to the different distribution of gravity anomaly features in different regions, based on the gravity anomaly data, K-means clustering method is used to cluster gravity anomaly features in different regions, so as to classify relatively similar gravity data into one category, divide the region where such gravity data is located into one category, and set the label of this region.

3.1 K-means clustering

This section uses K-means clustering gravity anomaly data for subclass classification. There are many methods for cluster analysis, and the commonly used measurement methods are as follows: using distance to measure the degree of similarity between samples. The smaller the distance between two samples, the higher the similarity between them. The greater the distance, the smaller the similarity.

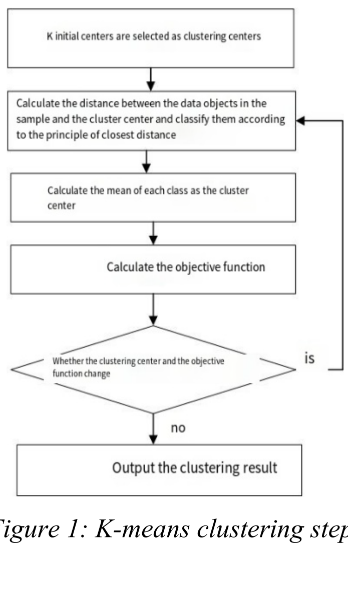


Figure 1: K-means clustering steps

ISSN 2522-3488 Vol. 7, Issue 9: 42-47, DOI: 10.25236/IJNDES.2023.070908

K-means clustering is a simple iterative clustering algorithm, and its calculation process is shown in Figure 1. Then, Euclidean distance is selected as the method to measure the similarity, and the contour coefficient method is used to determine the rationality of distance measurement.

Euclidean distance is defined as follows: Suppose there are two sample sums whose feature vectors are represented by and, then the distance of sample sums is:$x_i x_j (x_{i1}, x_{i2}, .. x_{ik},)(x_{j1}, x_{j2}, .. x_{jk},) x_i x_j$

$$D_{ij} = [\sum_{m=1}^{k} (x_{im} - x_{jm})^2]^{1/2} \tag{1}$$

The absolute distance is defined as follows: Suppose there are two sample sums, whose eigenvectors are represented by and respectively, then the distance of the sample sum is:$x_i x_j (x_{i1}, x_{i2}, .. x_{ik},)(x_{j1}, x_{j2}, .. x_{jk},) x_i x_j$

$$D_{ij} = \sum_{m=1}^{k} |x_{im} - x_{jm}| \tag{2}$$

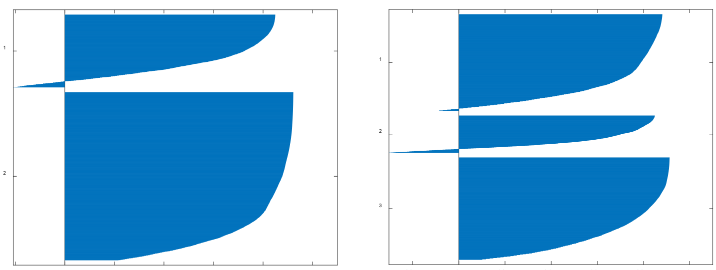
As can be seen from Figure 1, the selection of cluster center, the measurement method of distance between data object and cluster center will affect clustering

The accuracy of the result. Because the classification sample is 14883, the number of classification is not easy to be too much, so the number of cluster centers is selected as 8, 9 and 10, and the Euclidean distance is selected for cluster analysis calculation. The calculation results are compared with the contour coefficient method to finally determine the number of cluster centers.

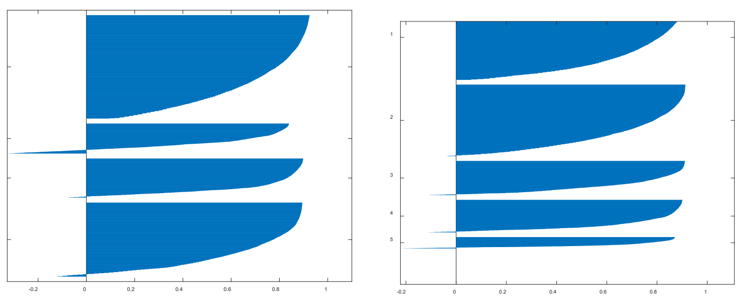
The contour coefficient is defined by the following formula, which represents the average distance between sample $a(i)$i and other samples of the same cluster, and represents the average distance between sample $b(i)$i and all samples of other clusters. When the calculated result is close to 1, it indicates that the classification $S(i)$of sample i tends to be reasonable; $S(i)$Close to -1, it means that the classification of sample i tends to be unreasonable, and it should be divided into other clusters. If it is close to 0, it means that sample i is on the boundary of two clusters.$S(i)$

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} = \begin{cases} 1 - \frac{a(i)}{b(i)} & a(i) < b(i) \\ 0 & a(i) = b(i) \\ \frac{b(i)}{a(i)} - 1 & a(i) > b(i) \end{cases} \tag{3}$$

Using 14883 sets of gravity anomaly data for clustering, FIG. 2 shows the distribution of contour coefficients when the number of selected cluster centers ranges from 2 to 11. It can be seen from the contour diagram that in 10 clustering cases, most points in each cluster have large contour values (greater than 0.8), indicating that these points can be well distinguished from neighboring clusters. However, there are also some points in each cluster that have lower contour values, indicating that they are close to points in other clusters.

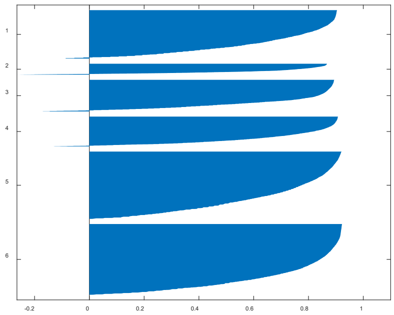


(a) The number of categories is 2 *(b) the number of categories is 3*

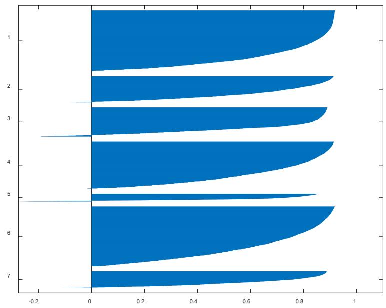


(c) The number of categories is 4 *(d) The number of categories is 5*

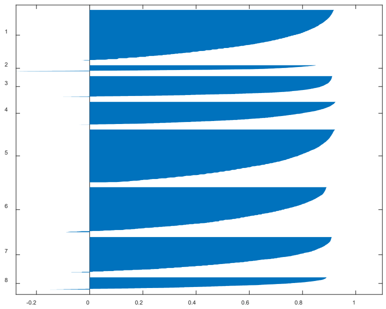
ISSN 2522-3488 Vol. 7, Issue 9: 42-47, DOI: 10.25236/IJNDES.2023.070908



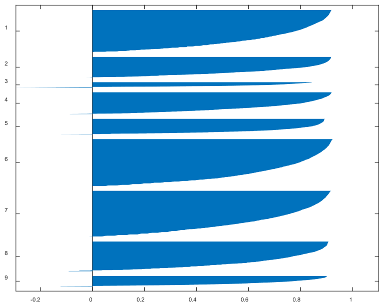
(e) The number of categories is 6



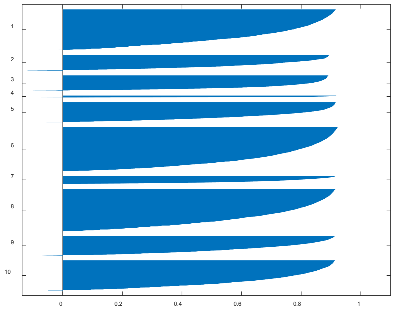
(f) The number of categories is 7



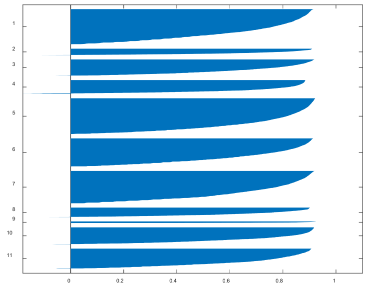
(g) The number of categories is 8



(h) The number of categories is 9



(i) The number of categories is 10



(j) The number of categories is 11

Figure 2: Distribution of contour coefficients with the number of cluster centers being 2-11

Set the number of iterations to 100, and at each iteration, the K-means algorithm redistributes points between clusters to reduce the sum of the distances from the points to the center of mass, and then recalculates the center of mass assigned by the new cluster. The sum of distances and the number of reassignments decrease with each iteration until the algorithm reaches a minimum value. The algorithm used in K-means consists of two stages. In all three of these clustering cases, the algorithm converges, and the second stage of the algorithm does not undergo any redistribution, indicating that the first stage reaches its minimum after a few iterations.

By default, K-means begins the clustering process with a randomly selected set of initial centroid locations. The K-means algorithm converges to the solution as a local minimum; That is, K-means can partition the data such that moving any point to another cluster increases the sum of distances. However, as with many other types of numerical minimization problems, the solution reached by K-means sometimes depends on the starting point. Therefore, there may be other solutions (local minimums) for this data that have a smaller sum of distances.

3.2 Region Fit calibration

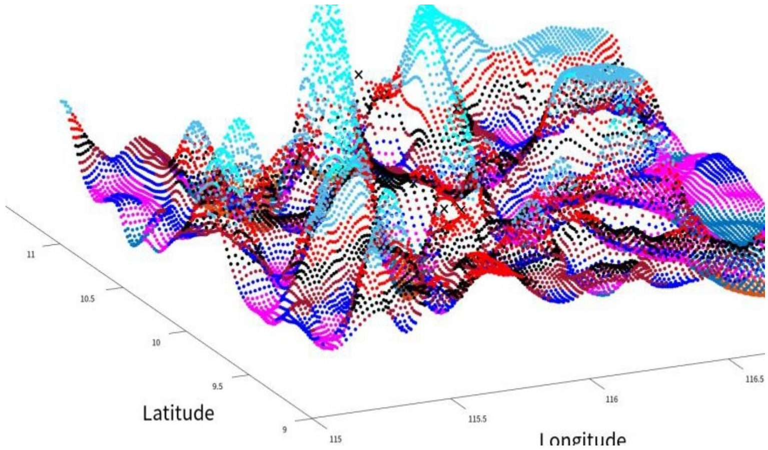
When the number of clustering centers is 2-11, the iteration results are shown in Table 1. It can be seen from the table that when the number of cluster centers is 2, 3, 5, 7, 8, 9, all 100 iterations converge, and the clustering effect is good; Combined with Figure 2 and the sum of the best distance, when the number of clusters is 2, 3 and 5, the sum of the best distance is larger and the distribution of contour coefficients has a negative extension; When the number of clusters is 7 and 8, the profile coefficient distribution has negative extension; When the number of clusters is 9, the contour coefficient is large and

ISSN 2522-3488 Vol. 7, Issue 9: 42-47, DOI: 10.25236/IJNDES.2023.070908

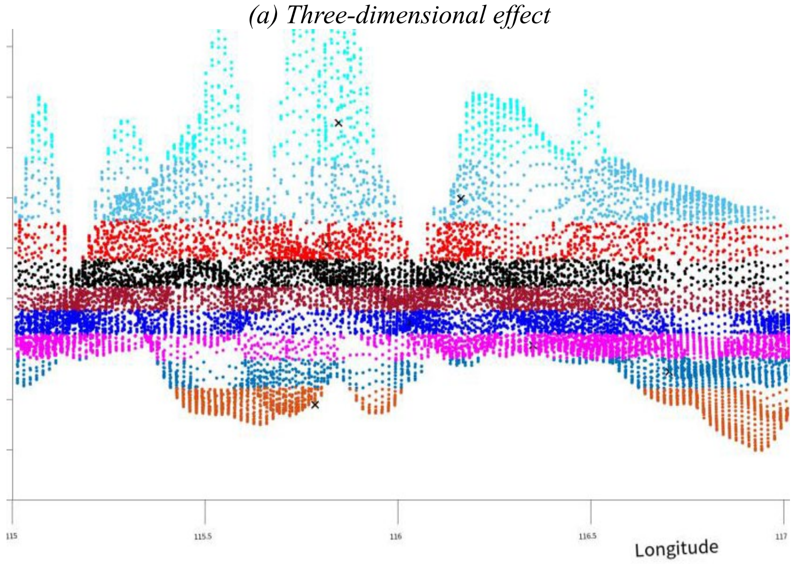
evenly distributed on the positive side, the part of the contour coefficient less than 0 is very small, and it converges during iteration, and the sum of the best distance is relatively small. Therefore, when the number of clusters is 9, the fit of each region is the highest. Figure. 3 shows the classification effect when the Euclidean distance is measured and the number of cluster centers is 9. As can be seen from the two-dimensional figure, the classification of each region is obvious, and it is classified layer by layer along with the altitude.

Table 1: Results of clustering iteration

Number of clusters	Sum of best distances	Convergence or not during iteration
2	205735	is
3	149831	is
4	122634	no
5	102476	is
6	87761.5	no
7	76790.3	is
8	69800.4	is
9	63711.2	is
10	58766.9	no
11	54869.2	no



(a) Three-dimensional effect



(b) Two-dimensional effect

Figure 3: Classification effect

According to the above analysis, the cluster labels of each group of data are obtained. From the results, it can be seen that when a total of 9 regions are divided, the region division has the best adaptability. Among them, subclass 1 is labeled as region 1 and labeled as 1; Label subclass 2 as area 2, label 2, and so on. The zone label results are saved, where some of the results are shown in Table 2.

Table 2: Zone label results

Data Sets	Longitude	Latitude	Gravitational outliers	Zone Tags
1	115.0083	11.0068	59.3	6
2	115.025	11.0068	58.1	6
3	115.0417	11.0068	52.5	6
4	115.0583	11.0068	45.5	1
5	115.075	11.0068	38.4	1
6	115.0917	11.0068	31.3	3
7	115.1083	11.0068	24.3	3
8	115.125	11.0068	18.2	7
9	115.1417	11.0068	13.4	2
10	115.1583	11.0068	10.5	2
11	115.175	11.0068	9.2	2
12	115.1917	11.0068	9.2	2
13	115.2083	11.0068	10.5	2
14	115.225	11.0068	13	2
15	115.2417	11.0068	16.3	7
16	115.2583	11.0068	19.5	7
17	115.275	11.0068	22	7
18	115.2917	11.0068	22.7	7
19	115.3083	11.0068	20.8	7
20	115.325	11.0068	16.6	7

4. Conclusion

In this paper, the underwater adaptive regions are classified and predicted by data-driven method, and the gravity anomalies in different regions are clustered by K-means clustering method. The experimental results show that when the number of clusters is 9, the fitting effect of each region is the best. The method proposed in this paper has a good application prospect in the field of underwater navigation and can provide strong support for adaptive region classification and prediction in practical application. Future research can further explore other clustering methods and optimization algorithms to improve the accuracy and practicability of underwater navigation.

References

[1] Li Zhenxing, Han Lina, Shi Nan. Research on Movie box Office Prediction based on Bayesian Classification Model [J]. Computer and Digital Engineering, 2020, 48(09):2233-2237.

[2]Chen Si. Research on Outsourcing Scheme of Naive Bayes Security Classification in Cloud Computing environment [J]. Computer Applications and Software, 2020, 37(07):275-280. (in Chinese)

[3] Jie, Tian, Xue Shan-hua, et al. Classification of underwater still objects based on multi-field features and SVM[J].Journal of Marine Science & Application, 2007.DOI:10.1007/s11804-007-6042-4.

[4] Lai Xiaofeng. News based on naive bayesian classification research [D]. Jiangxi university of finance and economics, 2020.

[5] TENG Tao, Wang Guojun, Zhou Weisheng et al. Research and application of GA-BP neural network model in classification and prediction of rockburst intensity [J]. Modern Mining, 2023, 39(09):278-281.

[6] Yang Yonghong, Wang Chun, Yang Chao et al. Risk identification of secondary highway alignment Accidents based on BP neural network [J/OL]. Journal of Shenzhen University (Science and Technology Edition):1-8[2023-11-04].

[7] Dou Yihua, Zhang Jiaqiang, Li Guoliang et al. To optimize the BP neural network based continuous tube fatigue life prediction [J]. Journal of petroleum machinery, ploidy of 2023 (10): 144-149.