

Adaptive region prediction of gravity aided navigation system based on SVM multi-classification and mixed Gaussian clustering model

Yarui Fu

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

Keywords: Gravity aided navigation; SVM multi-classification; Kmeans clustering; Gaussian mixed clustering; Adaptive region prediction

Abstract: This paper aims to solve the problem of determining the adaptive region in gravity aided navigation system. In order to achieve this goal, a SVM-based multi-classification method is proposed to determine the adaptive region. First, the data is divided into three dimensions and the applicability of the regions is determined based on the standard deviation. Secondly, the Kmeans clustering model and Gaussian mixed clustering model are established for comparison and analysis, and the optimal number of regions is 5. By observing and comparing the regions divided by the two methods, it is found that the regions divided by the Kmeans method are not continuous and relatively discrete, while the standard deviation of each region of the Gaussian mixed model is better than that of the Kmeans method. The standard deviations of the five regions divided by the Gaussian method are 17.12, 24.34, 26.28, 13.39, and 21.08, respectively. The corresponding regions are labeled 1-5. Zones 2, 3, and 5 are adaptation zones.

1. Introduction

In a gravity-assisted navigation system, in order to ensure the reliability and accuracy of navigation, it is necessary to select suitable underwater navigation areas, which are called "adaptation zones". Selecting the adaptive zones is a key step, but one of the most challenging problems is how to determine these adaptive zones. Before selecting an adaptation area, it is necessary to interpolate and encrypt the gravity datum of the studied sea area, the most basic of which is the gravity anomaly datum [1-3].

Gravity anomalies refer to the uneven distribution of material density inside the Earth, resulting in a deviation between the actual observed gravity value of the Earth and the theoretical normal gravity value. After excluding the influence of various interfering factors, the change in gravity caused only by the uneven distribution of the density of the earth's matter is called the gravity anomaly. On Earth, the underground density distribution in different regions causes changes in the gravitational field, which can be detected and recorded using gravity measurements [4].

In gravity-assisted navigation, observing changes in gravity anomalies is important for selecting an adaptation zone. If the gravity anomaly changes significantly in an area, then the navigation system

can achieve high-precision positioning in that area. Conversely, if the gravity anomaly in an area changes little, then the navigation system's positioning accuracy may not be too high [5, 6].

Because the distribution of gravity anomaly features varies from region to region, it is critical to build a classification model that can predict the adaptation region. These models can help determine where the navigation system can achieve high precision positioning and where it may face poor positioning accuracy. The selection of suitable area and gravity anomaly analysis is an important step to ensure the navigation accuracy of underwater vehicles.

This paper intends to solve the following problems. By refining the benchmark map, reasonably dividing regions for demonstration, and completing the adaptation calibration (label Y) of each region.

2. Research Methods

(1) First, perform 3D data visualization on the reference data A of gravity anomaly and check the variation area of gravity anomaly through observation.

(2) Secondly, define the criteria for judging the significant change of gravity anomaly. Considering that we need to find the area with significant change of gravity anomaly, we take the standard deviation of the change of gravity anomaly in this area as the judging condition, which can perfectly show the difference between individuals and groups.

(3) It is necessary to reasonably divide regions according to regional gravity anomaly changes, and complete the adaptation calibration of each region (label Y). It is obvious that this is a problem of unsupervised learning (clustering). It is necessary to build a clustering model to solve the region division problem.

3. Results and Analysis

3.1 Divide data into regions reasonably and complete the adaptation calibration of each region

3.1.1 Problem preparation model

(1) Define the gravity anomaly change

First, we need to define what a gravitational anomaly is. Gravity anomaly refers to the change of gravity caused only by the uneven distribution of the density of the earth's matter after excluding the influence of various interference factors. We can think that the magnitude of the standard deviation of the gravity anomaly represents the significance of the change of the gravity anomaly. Specifically, we can define:

$$s = \begin{cases} 1 & \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 > m \\ 0 & \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 \leq m \end{cases} \quad (1)$$

Where m is a threshold, such as 20, indicating a change in gravity anomalies. N Represents the amount of gravity anomaly data in the region, and represents the average value of gravity anomalies in the region \bar{x} .

(2) Data display

We can obtain the latitude and longitude and gravity outliers of the given location.

A 3D graph is used to visualize these data so that we can observe the variation of gravity anomalies at each location.

(3) Determine the significant variation of gravity anomalies in each region

First, calculate the standard deviation of gravity anomaly change in this region, and then judge whether significant gravity anomaly change occurs according to the threshold value.

3.2 Model Establishment

3.2.1 Kmeans model building

Find a partition scheme of K clusters, so that the loss function corresponding to the clustering result is minimum. K-means clustering is the most basic and commonly used clustering algorithm. Its basic idea is that through iterative finding, the loss function can be defined as the sum of error squares of each sample distance from the center point of the cluster to which it belongs:

$$J(c, \mu) = \sum_{i=1}^M \|x_i - \mu_{c_i}\|^2 \quad (2)$$

Where x_i represents the first sample, c_i is the cluster to which it belongs, μ_{c_i} represents the center point corresponding to the cluster, M is the total number of samples.

3.2.2 Establishment of Gaussian mixture clustering model

(1) Gaussian distribution probability density function:

The probability density function of Gaussian distribution is expressed as:

$$f(x; \mu, \Sigma) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right) \quad (3)$$

Where, x is the observed data point, μ is the mean of the Gaussian distribution, Σ is the covariance matrix, and d is the data dimension.

(2) Probability density function of mixed Gaussian distribution:

The probability density function of GMM is expressed as a linear combination of multiple Gaussian distributions:

$$f(x) = \sum_{i=1}^K \pi_i \cdot f(x; \mu_i, \Sigma_i) \quad (4)$$

Where, K is the number of mixed components, π_i is the mixing coefficient of each mixed component, and μ_i, Σ_i is the mean and covariance matrix of each mixed component.

(3) Expectation step of EM algorithm:

In step E, calculate the posterior probability (responsibility) that the data point belongs to the fourth mixed component:

$$w_i^{(t)} = \frac{\pi_i^{(t)} \cdot f(x; \mu_i^{(t)}, \Sigma_i^{(t)})}{\sum_{j=1}^K \pi_j^{(t)} \cdot f(x; \mu_j^{(t)}, \Sigma_j^{(t)})} \quad (5)$$

Where, $w_i^{(t)}$ indicating the posterior probability that the data point belongs to the fourth mixed component, at the first iteration, $\pi_i^{(t)}, \mu_i^{(t)}, \Sigma_i^{(t)}$ are the mixing coefficient, mean, and covariance matrix of the first mixed component.

(4) Maximization steps in EM algorithm:

In the M step, update the parameters of the mixed components to maximize the likelihood function:

$$\pi_i^{(t+1)} = \frac{1}{N} \sum_{j=1}^N w_j^{(t)} \quad (6)$$

$$\mu_i^{(t+1)} = \frac{\sum_{j=1}^N w_j^{(t)} \cdot x_j}{\sum_{j=1}^N w_j^{(t)}} \quad (7)$$

$$\Sigma_i^{(t+1)} = \frac{\sum_{j=1}^N w_j^{(t)} \cdot (x_j - \mu_i^{(t+1)}) \cdot (x_j - \mu_i^{(t+1)})^T}{\sum_{j=1}^N w_j^{(t)}} \quad (8)$$

Where, N is the total number of data points, x_j is the data point, $\pi_i^{(t+1)}$ is the posterior probability calculated in step E, and $\mu_i^{(t+1)}$ is the new mixing coefficient, mean and covariance matrix for the first mixing component i .

3.3 Model solution and comparison

3.3.1 Data display

First of all, the data presentation is carried out by reading the data from the Gravity anomaly reference data A file, normalizing the gravity anomaly value, scaling it to between 0 and 1, so that it can be used as the basis for color mapping.

Then create the color map and finally create a 3D scatterplot. In it, longitude and latitude determine the position of the data points, while the gravity outlier determines the height of the points. Set the color mapping so that the color of the scatter is based on the gravity outlier. This way, different data points will be rendered in different colors, representing the relative sizes of their gravitational outliers.

Figure 1 shows the gravity outlier data, with red indicating large gravity outliers and blue indicating small gravity outliers. From the figure, it can be simply analyzed that the gravity anomaly fluctuates significantly, with four distinct peaks and five sub-peaks.

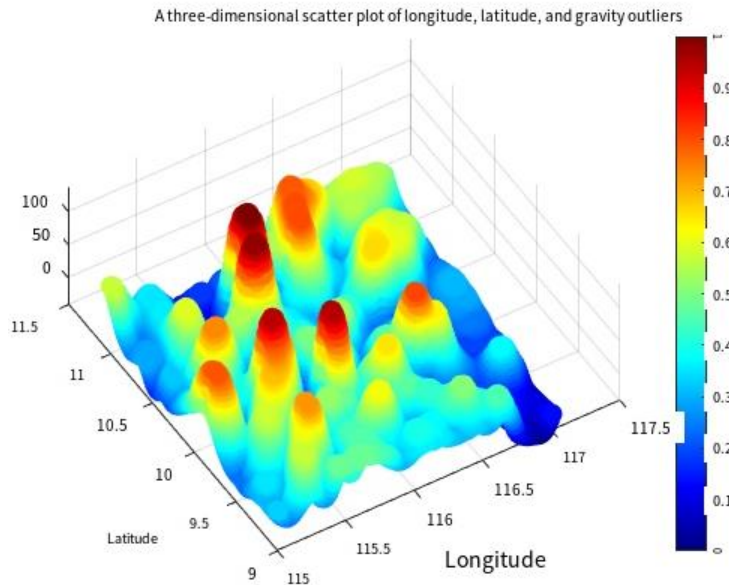


Figure 1: Three-dimensional scatterplot of gravity anomalies

3.3.2 KMeans model solution

(1) Randomly select K centers, denoted as $\mu_1^{(0)}, \mu_2^{(0)}, \dots, \mu_k^{(0)}$

(2) Define the loss function: $J(c, \mu) = \min \sum_{i=1}^M \|x_i - \mu_{c_i}\|^2$

(3) Let $t=0,1,2,\dots$ Is the number of iterative steps, repeat the following process to know convergence: J

(3.1) For each sample, assign it to the nearest center x_i

$$c_i^t < -\arg \min_k \|x_i - \mu_k^t\|^2 \quad (9)$$

(3.2) For each class center k, recalculate the center of the class

$$\mu_k^{(t+1)} < -\arg \min_{\mu} \sum_{i:c_i^t=k} \|x_i - \mu\|^2 \quad (10)$$

(a) Select the number of clusters K: First of all, regionalization into a class has the best effect according to several experiments. $k = 5$

(b) Initialize cluster center point: Randomly select K data points as the initial cluster center point, or use other initialization methods. These points will become the center of each cluster.

(c) Assign data points to clusters: For each data point, calculate the distance between it and the center point of each cluster, usually using Euclidean distance or another distance metric. The data point will be assigned to the closest cluster.

(d) Update cluster center points: Calculate the average of all the data points in each cluster, and then use these averages as the new cluster center points.

(e) Repeat allocation and update: Repeat steps 3 and 4 until the cluster center point no longer changes significantly, or a predetermined number of iterations is reached.

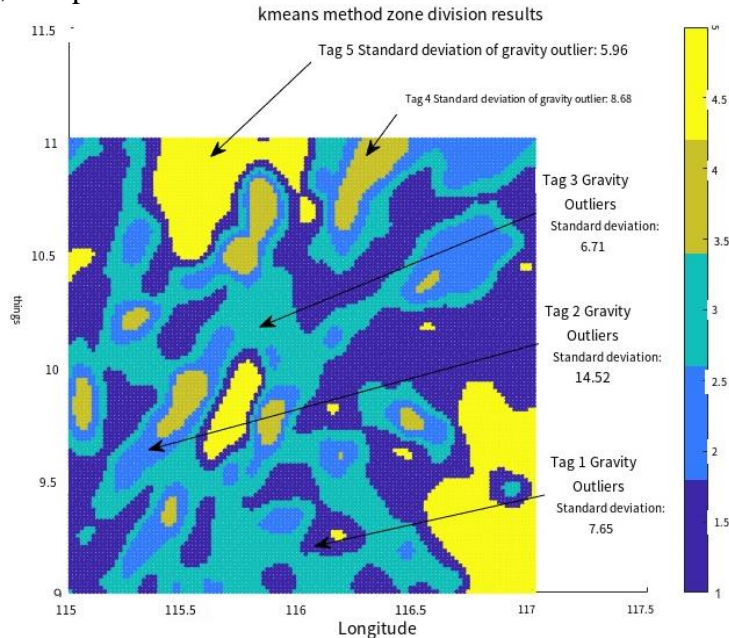


Figure 2: Result of kmeans region partitioning

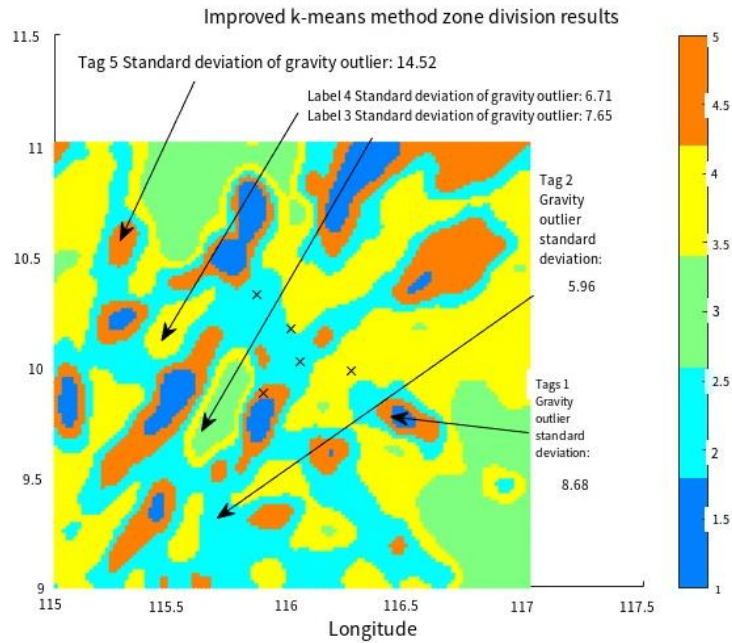


Figure 3: Region division results of kmeans method with improved parameters

It can be seen from the Figures 2 and 3 that the regions separated by conventional kmeans classification are relatively discrete, and the standard deviation of gravity outliers is not large, and most of them are concentrated in 5-8. In this paper, the parameters of kmean are adjusted, and it is found that the effect is not very good through the experiment again. However, after several experiments, it is determined that the best effect is to divide the region into five categories.

3.3.2 Gaussian mixture clustering model solution

GMM steps are as follows.

(1) Initialization: Randomly initialize the mean, covariance matrix and mixing coefficient of each Gaussian distribution. These parameters will be updated in subsequent iterations.

(2) Expectation Step (Step E): In this step, the probability (posterior probability) that it comes from each Gaussian distribution is calculated for each data point. This can be calculated using the Bayesian formula, which is the observed value of a given data point, calculating the probability that it belongs to each Gaussian distribution.

(3) Maximization step (Step M): In this step, the parameters of each Gaussian distribution are updated based on the posterior probability calculated in step E. Specifically, update the mean, covariance matrix, and mixing coefficients for each component to maximize the likelihood function.

(4) Iteration: Repeat step E and step M until convergence, usually when the parameter change is less than a certain threshold or reaches the maximum number of iterations.

(5) Result: After convergence, GMM produces a mixed model that fits the data, including parameters for each component. These components can be used to identify subgroups in the data or used to generate new data points.

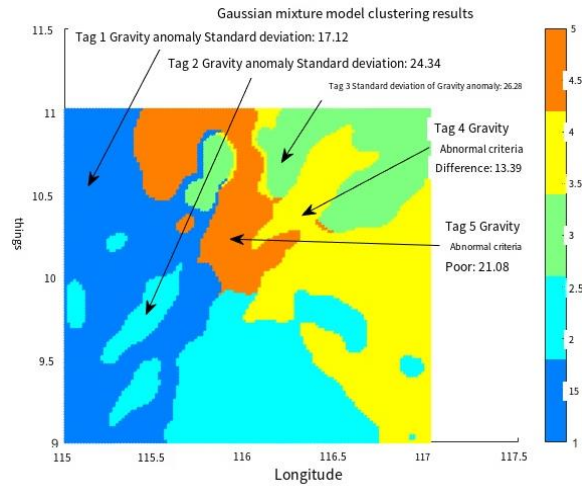


Figure 4: Clustering results of Gaussian mixture model

Previous experiments have determined that it is best to divide regions into five categories. The following work in this paper continues to divide regions into five categories. It can be seen from Figure 4 that the standard deviation of gravity of regions 2.3.5 reaches 20, label 2.3.5 is all adaptive regions, and the divided regions are relatively continuous and integral, which is also convenient for the navigation work of underwater vehicles in the background of the problem. In this paper, the labels after clustering are stored. The following is part of the results (Table 1).

Table 1: Cluster labels are displayed

Longitude	Latitude	Gravitational outliers	Tag y
115.0083	11.0068	59.3	1
115.025	11.0068	58.1	1
115.0417	11.0068	52.5	1
115.0583	11.0068	45.5	1
115.075	10.9904	36	1
115.1583	10.9904	12.9	1
115.2917	10.9904	22.2	1
115.3083	10.9904	20.6	1
115.325	10.9577	16.7	1
115.3417	10.9577	11.4	1
115.3583	10.9577	6.4	1
115.3083	10.794	18.2	1
115.325	10.794	12.6	1
115.325	9.7937	12	1
115.6417	9.7937	8.9	1
115.5583	9.0045	31.8	1
115.2917	10.2533	74.2	2
115.3083	10.2533	75.5	2
115.325	10.2533	74.6	2
115.2417	10.204	79.3	2
115.2583	10.204	86.3	2
115.275	10.204	90.1	2
115.2417	10.1877	78.1	2
115.2583	10.1877	83.6	2
115.275	10.1877	86	2
115.625	10.04	57.2	2
115.6083	9.9579	67.7	2
115.475	9.9415	84.5	2

115.6083	9.9415	71	2
116.7083	9.0045	0.6	2
116.725	9.0045	8.9	2
116.2083	11.0068	41.4	3
116.225	11.0068	48.4	3
116.2417	11.0068	58.1	3
116.375	10.9904	83.7	3
116.3917	10.9904	81.4	3
116.4083	10.9904	79.9	3
116.975	10.9904	48.2	3
116.3417	10.9577	91.2	3
116.375	10.9577	85.1	3
116.4083	10.9577	80	3
116.8583	10.9413	55.3	3
116.7083	10.9249	56.1	3
116.725	10.9249	55.5	3
116.9417	10.9249	39.4	3
116.9917	10.9086	20.2	3
116.0917	10.8922	38.1	3
116.5917	10.5155	65.6	3
115.675	10.4664	76.5	3
116.8417	10.4664	45.4	3
116.5417	10.2696	27.5	3
116.0583	11.0068	14.7	4
116.075	11.0068	26.7	4
116.0917	11.0068	35.6	4
116.125	10.9577	32.3	4
116.1417	10.9577	32.8	4
116.1083	10.9413	31	4
116.0583	10.9086	17.6	4
116.0583	10.8922	22.3	4
116.425	10.8104	18.8	4
116.525	10.0892	1.3	4
116.6083	10.0892	0.4	4
116.0917	9.7609	4.9	4
117.0083	9.6623	7.2	4
116.9583	9.6459	10.8	4
117.0083	9.6459	7.1	4
117.0083	9.0045	17.6	4
115.4417	11.0068	13.5	5
115.4583	11.0068	17.5	5
115.475	11.0068	19.6	5
115.4917	11.0068	- 19	5
115.5083	11.0068	17.1	5
116.225	10.2533	26.9	5
116.2417	10.2533	29.5	5
115.8417	10.2205	27	5
115.9083	10.2205	28.1	5
116.0583	10.1056	25.4	5
116.0417	10.0564	25.8	5
116.0583	10.04	28.6	5
116.025	9.8922	33.4	5

4. Conclusions

In this paper, the adaptive region of gravity aided navigation system is predicted by the establishment of SVM multi-classification and mixed Gaussian clustering model. The results show that the Gaussian mixed cluster model is superior to the Kmeans cluster model in both continuity and standard deviation of regional division. Through the application of the model, the adaptive region can be predicted effectively, which provides an important reference for the navigation accuracy of the gravity aided navigation system in different regions. In addition, the research method of this paper can also provide reference for the solution of similar problems.

References

- [1] F. P. Wang, M. H. Yang, D. Q. Xia et al. *Research on n/γ screening method of liquid flash detector based on small-batch KMEans++clustering algorithm [J]. Nuclear Electronics and Detection Technology*, 2022, 42(06):1073-1081.
- [2] B. S. Li, C. D. Qin. *Research on Price prediction of electric vehicle based on SVM multi-classification based on particle swarm optimization [J]. Computer Science*, 2019, 47(S2):421-424.
- [3] W. H. Cui, S. Zheng, S. Q. Yang et al. *Gaussian Mixture model for nuclear power operation condition classification based on density peak clustering [J]. Science Technology and Engineering*, 2023, 23(20):8670-8676.
- [4] X. Liu, H. W. ZHENG. *Construction of Cuiguan Pear size grade evaluation Model based on support vector Machine [J]. Science and Technology Innovation*, 2023(24):221-224.
- [5] Wang B , Zhu J , Ma Z ,et al.*Improved Particle Filter-Based Matching Method With Gravity Sample Vector for Underwater Gravity-Aided Navigation[J].IEEE Transactions on Industrial Electronics*, 2020, PP(99):1-1.DOI:10.1109/TIE.2020.2988227.
- [6] Pan Shidong, Zhang Hui. *Highway land use prediction based on multiple linear regression model: A case study of Xuancheng City [J]. Green science and technology*, 2023, 25 (4): 235-239+271.