Caspian Journal of Applied Mathematics, Ecology and Economics
V. 7, No 1, 2019, July
ISSN 1560-4055

Mathematical Approaches to Ground Objects Classification According to Satellite Data

Z.C. Zabidov

Abstract. In modern times, many international organizations have been developing research methods for remote diagnostic of ground objects. Over the last 15-20 years, despite the wide-scale development of computer programs that allow space designs to be of qualitatively new materials (up to 0.5 m accuracy) and to process cosmic images, the problem of using satellite imagery for ground objects classification has not been solved practically.
Comparative mathematical approaches to solving the land classification of satellite data, and comparative mathematical approaches to solution are given. Satisfaction with the application of satellite classification according to the satellite data of the classification and recognition methods.

Key Words and Phrases: satellite data, classification, metric distance, cluster algorithms.

2010 Mathematics Subject Classifications: 94A12

---

1. Introduction

Rapid development of computer science and the broad range of software systems (e.g. Matlab, Mathematics, Mapple, etc.) have enabled the satellite data to be used for different authentication and recognition issues, or stimulates the accuracy of solving algorithms. The accuracy of the problem solved depends on many factors. Let's note some of them.

1) The mathematical specificity of the problem, in other words, the mathematical model's realistic relevance, so certain assumptions (restrictions) are taken for the model's probability, which in turn creates possible errors;

2) Mathematical problems typically have inverse issues in nature, and these issues are non-corrupt;

3) The solution of the problem is based on statistical data, the prices of these data are coincidental and vary depending on many factors; Statistical data is insufficient;

4) Based on the solution of the problem, the division of the classroom is broken and the classification criteria are different, and finally the solution depends on the chosen method; and so on.

In general, the task of recognition is as follows. The most common definition of a class is the following: a class is a collection (family) of objects that have some common

 © *2013 CJAMEE* All rights reserved.

properties. Information about the properties of an object can be obtained by observations, measurements, assessments, etc. and represented by a set of features, the values of which are expressed in numerical scales. Objects belonging to the same class are considered indistinguishable (equivalent), and each class of objects is characterized by a certain quality that distinguishes it from other classes. Together, all classes must constitute the initial set of objects [1].

Spectral images of objects on the earth's surface are non-stationary, as they depend on many factors, such as topography, soil type, climate, geographical location. To increase the reliability of decisions, it is necessary to use a priori information about the geometry of the survey, on the one hand, and the contextual information of the images themselves - on the other.

For this reason, the application of new approaches and the comparative analysis of the obtained results with known outcomes are of great importance both from the theoretical and the point of view.

2. Methodical basis and calculation methodology

The offered work is devoted to the classification of ground objects (e.g. of soil types, of aerosol-gas compounds) according to satellite data. The issue is as follows.

The classification of objects of the given area D is known in the coordinates (points) $\{P_i : i = \overline{1,m}\} \subset D$. Let us point the known classification of objects, in other words, objects classes with $\{M_k : k = \overline{1,r}\} : M_i \cap M_j = \emptyset, i \neq j$. Thus for $\forall i \in 1 : m = \{1; 2; ...; m\}$, $\exists k\,(i) \in 1 : r$

$$P_i \subset M_{k(i)},$$

it is true. In addition, at each $P \in D$ point the values of $W_{\lambda_k}(P), k = \overline{1,\chi}$- satellite data are known in the wave lengths $\{\lambda_k : k = \overline{1,\chi}\}$. The problem is to find the class M_k, of $\forall P \in D$ point, in other words, to find $\exists k\,(p) \in 1 : r$ number for $\forall P \in D$, so that $P \in M_{k(p)}$ is true.

First of all, it should be noted that, according to satellite data, the issue of ground objects classification has particular peculiarities. Thus, the variability of ground objects cover (eg: vegetation - natural or artificial, snow cover, other artificial covers, etc.) causes the variability of satellite data (at the same time). This, in turn, contributes to the distortion of the result. Thus, according to the satellite data, the reflection of object cover is first identified, and in the next step it is necessary to analyze the relationship between the value of reflection and object classification.

Now let's look to the problem solving algorithm.

Let's define the following set of indexes.

$$I_k = \{i \in 1 : m / P_i \in M_k\}, k = \overline{1,r}.$$

It is obvious that,

$$I_i \bigcap I_j = \emptyset, i \neq j$$

is true. The accuracy of the solution depends, of course, on the fact that the measurement data, i.e. enough statistical data.

Another peculiarity of the problem is that statistics are not usually sufficient (e.g. in large areas or in inaccessible areas). In this case, there is a need for new and deeper analysis methods to increase accuracy.

Each $P \in D$ point corresponds to the vector $P\,(\omega; W_1; ...; W_\kappa)$ where ω is the ground class of point P, where $W_k, k = \overline{1,\kappa}$ - is satellite datas. In case of solving the problem, each vector $(W_1; ...; W_\kappa)$ corresponds to a class. Let's point this argument with π, i.e. $\pi\,(W_1; ...; W_\kappa) = \omega$. Thus, the mathematical implication of the problem consists in the construction of $\pi : R^\kappa \to \left\{M_k : k = \overline{1,r}\right\}$. The values of the π judgment are sets, i.e. clusters. In general, the problem does not have this kind of single solution. Since the solution process is based on statistical data, the value of π is a coincidental number, for each $\overrightarrow{W} = (W_1; ...; W_\kappa)$ vector $\pi\left(\overrightarrow{W}\right)$ is a random quantity, i.e. $\pi\left(\overrightarrow{W}\right) \in M_k, k = \overline{1,r}$ occurs in a probability:

$$\sum_{k=1}^{r} P_k\left(\overrightarrow{W}\right) = 1.$$

Another approach to the study of $\pi\,(\bullet)$ is the application of phases theory methods. In this case, the value of $\pi\left(\overrightarrow{W}\right)$ can belong to each M_k class by defining an affiliation function. We will use Appendix I and II for the determination of $\pi\,(\bullet)$ in this case.

I. In this case, it is assumed that the classes are separated by satellite data, otherwise, there is no single solution to the problem. The center of each class is found by:

$$\overrightarrow{W}_0^{(k)} = \frac{1}{|I_k|}\sum_{i \in I_k} \overrightarrow{W}_i, k \in 1 : \kappa$$

here $|I_k|$ - the number of elements in I_k.

Should be find $R_k > 0$ radius, that $O_{R_k}\left(\overrightarrow{W}_0^{(k)}\right)$ ($\overrightarrow{W}_0^{(k)}$center, R_k radius) balls

$$O_{R_i}\left(\overrightarrow{W}_0^{(i)}\right) \bigcap O_{R_j}\left(\overrightarrow{W}_0^{(j)}\right) = \emptyset, i \neq j$$

satisfy the condition and $M_k \subset \min\limits_{R_k > 0} O_{R_k}\left(\overrightarrow{W}_0^{(k)}\right)$.

For random $\overrightarrow{W} \in R^\kappa$ ($\forall P\left(\omega; \overrightarrow{W}\right) \in D$ point) if there is $\exists k_0 \in 1 : \kappa$,

$$\overrightarrow{W} \in O_{R_{k_0}}\left(\overrightarrow{W}_0^{(k_0)}\right)$$

then the corresponding point belongs to the class M_{k_0}. Otherwise there is a need for further analysis. For example

$$k_0 = \min_{1 \le k \le \kappa}\left|\overrightarrow{W} - \overrightarrow{W_0}^{(k)}\right| \tag{1}$$

can be taken to $p \in M_{k_0}$. In the case of (1), if k_0 is uniquely determined value, there is still need for additional analysis.

The closeness of the point or the given $\overrightarrow{W}$ vector to any M_k class can be defined as a mean distance from this vector to the M_k class vectors, i.e.,

$$\rho\left(\overrightarrow{W}; M_k\right) = \frac{1}{|I_k|} \sum_{i \in I_k} \left\| \overrightarrow{W} - \overrightarrow{W_i} \right\|.$$

Then $\overrightarrow{W} \in M_{k_0}$, where

$$k_0 = \min_{k \in 1:\kappa} \rho\left(\overrightarrow{W}; M_k\right), \tag{2}$$

can be accepted. There is a need for further analysis of the case (2), which is uniquely in relation to k_0 number.

II. In this approach, the probability of the point belonging to a particular class can be determined by the degree of closeness to that class compared to all classes. Obviously, the closer the point to the class, the greater the probability of belonging to that class. Thus, the probability that the $\overrightarrow{W}$ vector belongs to the M_k class (P_k):

$$P_k = \frac{1}{r-1} \left(1 - \frac{\rho\left(\overrightarrow{W}; M_k\right)}{\sum_{i \in 1:r} \rho\left(\overrightarrow{W}; M_i\right)} \right), k \in 1:r, \tag{3}$$

can be calculated by formula. According to logic, if the $\overrightarrow{W}$ vector coincides with an element of any class, then the probability that this element belongs to that class must be "1". But according to the formula (3) it is not correct. Nevertheless, in the considered assumptions the probability of being the smallest distance from the point to its class and ultimately related to the class is greater.

Note that in the formulas (2) and (3), the probability of the distance is, for example, the dispersion of the difference of M_k with the $\overrightarrow{W}$ vector and so oncan be taken. In general, the distance is

$$\tilde{\rho}\left(\overrightarrow{W}; M_k\right) = \frac{1}{|I_k|} \left(\sum_{i \in I_k} \alpha_k \left\| \overrightarrow{W} - \overrightarrow{W_i} \right\|^{\rho} \right)^{1/\rho},$$

where $\sum_{\alpha_k \geq 0} \alpha_k = 1$ - weight coefficients, $p \in [1; +\infty)$ - numbers.

At the beginning of the article, the broad possibilities of various software systems (eg Matlab, Mathematics, Mapple, etc.) were noted for the possibility of using satellite data in different identification and recognition issues. On our part, the possibilities of applying the automatic classification of ground object class for satellite data of the classification and recognition methods included in the Matlab software system were investigated. The pdist, the linkage, and the cluster included in the MATLAB packet programs were used to perform calculations [2].

3. The results of calculations

In the present study [4], the remote sensing data shown in his / her scientific work is used. [4] uses a description of the September 13, 2006, satellite of Quickbird American satellite, located in Canibek, a geographical area of 49.35-49.43 ° and a geographical latitude of 46.75-46.84 .The images were taken by blue, green, red and NIR spectral bands. The coverage area is 65 km ^{2}. Surface measurements cover the years 2002-2009 and have been implemented by the author of the thesis. The total area of the survey is 50.6 km^2, of which 16.1 km^2 is the stationary.

As a result of the research, the author carried out the classification of the land at the 271 surface measurement point. It has been shown that there are mixed soil at 101 surface measurement points and no soil type has been identified at these measuring points. Land types have been identified at 172 surface measurement points. Land at the measurement points - A (black earth); B (chestnut soil), C (deserted salt); D (salty) soil types. At the point $P\left(\omega; W_1; ...; W_\kappa\right)$, the set of values ω is {A, B, C, D}. Here, the elements of the $\overrightarrow{W}$ vector are the indicator of the spectral reflection of the soil at the measuring point and the numerical values mentioned in the spectral channels (blue, green, red and NIR interval, MDVI Index). The number of points that the investigator points to different types of soil is given below.

$$I_A = 23, I_B = 45, I_C = 56, I_D = 45.$$

In the present case, using the metric distances and classification methods, the classification of the above described lands is carried out. The pdist, the linkage, and the cluster included in the MATLAB packet programs were used to perform calculations [2]. The number of points to different types was calculated using the possible variants of the metric distance and the classification algorithms. The metric distance used and the name of the classification algorithms and the results obtained are given in Table 2.

Table 2. Number of points on different types of soil
(according to the selected metric distance and classification algorithms)

The used metric distance and cluster algorithms	I_A	I_B	I_C	I_D
Euclidean distance and "nearest neighbor" algorithm	1	158	1	12
Euclidean distance and "remote neighborhood" algorithm	72	45	12	43
Euclidean distance and "medium" algorithm	37	112	11	12
Euclidean distance and "centralization" algorithm	55	49	12	56
Euclidean distance and "step-by-step" algorithm	37	106	17	12

Table 2 continuation

The used metric distance and cluster algorithms	I_A	I_B	I_C	I_D
Normalized Euclidean distance and ”nearest neighbor” algorithm	1	158	1	12
Normalized Euclidean distance and ”remote neighborhood” algorithm	2	92	12	66
Normalized Euclidean distance and ”medium” algorithm	39	107	14	12
Normalized Euclidean distance and ”centralization” algorithm	19	76	12	65
Normalized Euclidean distance and ”step-by-step” algorithm	38	108	14	12
Distance from the city neighbourhood and ”nearest neighbour” algorithm	1	168	2	1
Distance from city neighbourhood and ”remote neighbour” algorithm	38	86	39	9
Distance from city neighnourhood and ”mid-contact” algorithm	5	140	14	13
Distance from city neighnourhood and the ”centralization” algorithm	9	108	13	42
Distance from city neighnourhood and ”step-by-step” algorithm	2	154	3	13

Table 2 continuation

The used metric distance and cluster algorithms	I_A	I_B	I_C	I_D
Mahalanobis distance and "nearest neighbor" algorithm	1	158	1	12
Mahalanobis distance and "remote neighbor" algorithm	72	45	12	43
Mahalanobis distance and the "mid-contact" algorithm	37	112	11	12
Mahalanobis distance and the "centralization" algorithm	55	49	12	56
Mahalanobis distance and "step-by-step" algorithm	37	106	17	12
Distance in Minkowski metric and "nearest neighbor" algorithm	1	158	1	12
Distance in Minkowski metric and "remote neighbor" algorithm	12	16	23	121
Distance in Minkowski metric and the "mid-contact" algorithm	35	114	11	12
Distance in Minkowski metric and the "centralization" algorithm	55	48	12	57
Distance in Minkowski metric and "step-by-step" algorithm	37	112	11	12

It appears from the table that the results obtained from the classifications and methods used in calculations are not consistent with the type of soil types taken by the author of the case.

4. Conclusion

According to satellite data, classification of ground object classification is mathematically correct and there are different mathematical approaches to the solution. The use of standard clustering methods is not satisfactory for the classification of soil according to satellite data. The solution of the problem necessitates the application of non-standard approaches and the comparative analysis of the obtained results with known outcomes.

References

[1] S.I. Kolesnikova, *Methods for analyzing the informativeness of various types of attributes*, Tomsk State University Bulletin, 2009, No 1(6), pp. 69-80.

[2] N.V. Kolodnikova, *Review of textural attributes for pattern recognition tasks*, TUSUR reports, Automated systems for information processing, control and design, 2004, pp. 113-124.

[3] M.V. Konyushkova, *Mapping of the soil cover and soil salinity of the solonetz complex based on digital analysis of space imagery (using the example of the Dzhanybek stationary district)*, PhD dissertation, Moscow, 2010, pp. 155.

[4] V. Dyakonov, V. Kruglov, *Matlab Mathematical Expansion Packages*, Special handbook, 488 p.

Zakir C. Zabidov
Institute of Mathematics and Mechanics, NAS of Azerbaijan, Baku, Azerbaijan
Email: zakir_zabidov@mail.ru

Received: 05 February 2019
Accepted: 24 April 2019