THE REALIZATIONS OF THE GRASSLANDS CLASSIFICATION PRINCIPLES IN COMPUTER PROCESSING FOR THE GRASSLANDS RESOURCES INVESTIGATIONS

Shu Ning
Wuhan Technical University of Surveying and Mapping
Avenue Lo-Yu, 39 Wuchang
Wuhan, CHINA
Commission IV

Abstract

This paper discusses the classification principles about grass resources investigations, the possibilities of realizing those principles in computer processing, and the flexibilities of initially utilizing experts thinking. Because of the importance of auxiliary data in computer aided classifications procedure, this paper discusses also how to use those data, particularly the image data composition technologies for efficient employments. At last, it shows us a successful instance of computer aided classifications for the grasslands resources investigations with larger area and higher accuracy in Northern Tibet, the effective way for the automations in remote sensing applications and the correct direction of realizing experts thinking and making computer processing more intelligent.

At present, all of us have a consistant conclusion for the computer classifications using remote sensing data that it is insufficient to utilize only the spectral informations for some resource investigations and it can't get a satisfactory result. In this case, how to obtain the higher accuracy for the computer aided classifications? How to take into practice the expert's consciousness and approaches during classification procedure, just as the behaviours in visual interpretations? How to make auto-classifications more similar to the actions of specialists and remedy the insufficies of the visual interpretations? Now, we are going to discuss the realizations of some expert thinking which are flexible and easy to be carried out, and to introduce a successful experiment of grasslands resources investigations by means of computer aided classifications.

1. The principles of grasslands synthetic sequential classifications

Generally, the grasslands synthetic sequential classifications are divided into three "layers" or levels, which are classes, subclasses and groups. The first level is class, which is dependent to the climate conditions. One must take into account the water and heat indexes K and Q in this level, here K is humidity and Q the temperature sum in a year. Those two indexes are different with the varying of longitude and latitude, also the varying of hight. For example, the temperature falls 0.6 degrees (C) when the hight value incresses 100 m in the middle latitude regions, but humidities increase with the varying of elevation. In mountain regions, on the other hand, the gradient of temperatures and humidities on the southern and the northern slopes are not completely the same. So, because of the variations of temperature and humidity, the growing conditions of grass and the distributions of grasslands resources in different areas are not identical. The vertical distributions of grasslands are very distinct on the mountains and plateaus.

Generally, if K or Q is expressed by Y, we have

$$Y = f\,(L, A, H, O, S, R) \tag{1}$$

here, L is longitude, A is latitude, H represents elevation, O represents the orientation of mountain slope, S is slope, and R is the varying frequency of local relief. So,

$$DY = (\partial Y/\partial L)*DL + (\partial Y/\partial A)*DA + (\partial Y/\partial H)*DH + (\partial Y/\partial O)*DO + (\partial Y/\partial S)*DS + (\partial Y/\partial R)*DR \tag{2}$$

In practice, when calculating the gradient of K and Q, it is assessed according to the observation data at the meteorological stations which are located at different hight in the concerned region.

Subclasses are second level, which have the indexes of diverse relief and soil conditions, geomorphologic subareas and units, as well as the soil conditions must be taken into account. The usual geomorphologic subareas involve plains, mountains, hills, basins, valleys, etc. With the variations of geomorphologic and soil conditions, the distributions of grasslands are different. This kind of variation characteristics are mainly the horizontal district distributions.

The third level surveying is carried out by the determinations of groups, subgroups, microgroups of grasslands and the compounds of groups which are dependent on the vegetation conditions, and need the sampling analysis and the understanding of vegetation appearance characteristics and ground conditions, including superiority degrees, (i.e. cover percentage, hight, amount, weight, etc.) phenologies and living strengths. Those mean the very concrete indexes. When determining the grass superiority and the sub-superiority, one can name the groups. For example, the group Kobresia-Stipa purpurea means that Kobresia is the superiority grass, and the later is the subsuperiority grass.

Above is the summary of the principle that one must follow for the grasslands resources investigations, and that could be illustrated as fig. 1.

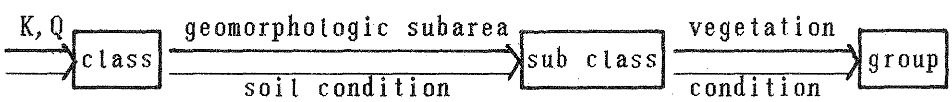

fig. 1

2. The realizations of the grasslands classification principles in the computer processing procedures

In the past computer aided classifications procedures, we usually take the multispectral data as objectives, and employ a certain kind of classifier. There exist a number of limitations when using this kind of method which purely depends upon the separation degree of different classes in the spectral space. It can't do any thing when there are different spectral reflections for a same kind of ground feature, or similar spectral reflection for different kinds of ground features.

According to the general reqularities of artificial intelligences, one must divide the train of the thoughts and the actions of some experts into several steps when implementing the human's thinkings and behaviours by a certain kind of machine, then realize those steps one by one. Can we do some things like that for the computer aided classifications of grasslands? This is a very interesting and a worthy problem for us. In accordance with our studies and experiments, it is possible to realize the grasslands classification principles in a certain extent.

The first, the water and heat indexes must be taken into consideration for the first level of classifications. In the mountain regions, those mainly concern the vertical distribution. When visually interpreting the photos, we can set up a classification system for that kind of distribution according to the stereoscopies (for example, the Spacelab image stereopairs.) and some reference documents. But how to take that into practice during the computer processing? It is necessary to use digital terrain modle (DTM) for the logical judgements in the procedure of classifications. It means that we can correctly determine some kind of grasslands which is growing in a certain elevation belt and has a certain range of spectral reflection. So we can get higher precision of classifications by combining the spectral informations with elevation informations (here, the first three terms of formula (2) are taken into account and DO, DS, DR are considered as zaro.), escaping from some negative functions of spectral informations. Furthermore, if considering the slopes and their orientations (i.e. consider the effects of DO, DS in formula (2).), those kind of data can be derived from DTM data. The informations concerning the slopes can also be employed in the logical processing.

Using DTM is only the indirect considerations about water and heat indexes. Generally, when considering directly those indexes, we can calculate the K value as following formula

$$K = \frac{r}{0.1 * Q} \tag{3}$$

(here r is the annual rainfall and unit is mm.)
and insert its variation gradient, in terms of the data from each meteorological observatory (or station) which are located at different hight in the concerning region. Like that, the curved surface of the water and heat indexes is set up as the reference curved surface and controlling paramenters for the grasslands resources investigations.

The geomorphologic conditions and soil condition must be taken into consideration for the second level of classifications. It is easier to be done when the synthetic analysis is in progress for the visual interpretations. Through the photo analysis, or refering to the concerning documents, or utilizing topographic maps, one can understand the geomorphologic situations. The soil conditions could be partly understood through the stereocopies, or by using soil theme maps, or by refering to some documents. But when computer works, one has to tell those informations to it, and train it. How to tell it? One method is segmentations, for example, covering all of the mountain informations, or removing that with the plain informations left only. The other is the coding in digit, i.e. inputing the geomorphologic subareas and units informations into computer in code form. As to the soil informations, it can be similarly obtained. In this case, computer "knows" what background or condition there is for some image informations. Then, the human thinking and logical judgements can be taken into practice by some programs to make computers as the "technologists", who can correctly classify the grass groups according to the human will in terms of the synthetic code concerning a certain geomorphologic and soil situations.

The considerations about vegetation conditions, the concrete appearance characteristics, growth state, living strength and ground conditions of various grass groups are taken into the third level of classifications. In terms of the results of field sampling and analysis, one can carry out inferences, judgements and drawing for the visual interpretations. But when using computer for the processing, it depends only upon the synthetic spectral informations offered by remote sensing, though it is not certain to depend directly upon some kind of those. The processing could base on the extracted (or enhanced) informations. Generally, the spectral information Is is the function of the factors such as the appearence characteristics, the growing and ground situations of grass groups, and image grey level data Gij (here i is line number, j is column.) is the function of the spectral informations Is of all of the ground features in a corresponding ground block of a pixel, which could be obtained in average value, or determined in terms of some sensor's parameters (for example, the optical characteristics of objective lens.) Because of the composite characteristics, the spectral informations become "blurred", and it had better to implement some enhancement processings firstly to form a new feature data set and then carry out the classifications. For this purpose one can implement some transformations, enhancements or local texture analysis for digital images, for example, the calculations of biomass index, the principle components transformations, the calculations of local grey level gradients, etc. We can also employ the technologies of image data compositions. For example, when some kind of remote sensing data are insufficient in spectrum for use, or its space resolution is not higher, one needs the use of another kind of data or the data with higher space resolution or spectral resolution for the utilizations. Those are all possible and flexible, and there is a instance that some one has reported a better effect obtained by using the composition of SPOT and TM image data while the China-France Congress on the Satellite Remote Sensing and the First Utilizations of

SPOT Image Data held in 1986. The combining of digital image enhancements and transformations technologies with composition technologies give us more and lerger space for the remote sensing data processing, and show us more possibilities to make the computer processing more intelligent.

In brief, it can be taken into practice to use the theories and consciousness of specialists during the computer aided classifications for grasslands investigations in accordance with the above analysis.

3. The utilizations of auxilary data

It can be found out from the above mentions that one of the important means is to utilize auxiliary data for the purpose to make the auto-classifications intelligent. For instance, the employment of DTM data can reflect the vertical distribution regularity of grasslands, and the horizontal distributions can be reflected by using the digitalizations of soil type distribution maps and geomorphologic unit distribution maps. In fact, the vizual interpretations of remote sensing images base upon a high dimension information space which is defined as the following:

$$S = \{ (l, p, X, Y, , Z, dm, an) \mid l, p \in Si;\ X, Y, Z \in St;\ dm(q) \in Dm,\ m = 1, 2, \ldots k,\ q = f(m);\ an \in An, n = 1, 2, \ldots r. \} \quad (4)$$

here, l,p are the coordinates (line, column) in image system Si, X,Y,Z the coordinates in terrain system St, Dm is the set of different kinds of image data (for example, the color infra-red photo, SPOT image, Landsat image, etc.), dm(q) represents the q band information of a certain kind data, (For example, d1(7) represents the seventh band information of Landsat image. It has different value for q to different image.) An represents the different kind of auxiliary data, including some reference documents and the ground sampling data. (For example, A1 is soil type, A2 is the agricultural division data, A3 is the agricultural meteorological data, etc.) While implementing the vizual interpretation, one must use so many informations for the synthetical analysis, the correlation analysis and the logical inferences, etc. Here, using An information is a very important factor. The computer processing in the past only uses a certain kind information of Dm (only one kind of image data in most case), and the "intelligent quotient" is not higher, so one must combine the Dm data and An data to make computer more intelligent, and let m, n larger than 1, if possible. But the problem now is how to use those auxiliary data and where we use that?

Generally, there are two existance mode of auxiliary data for the computer processing. One is direct employment with their value unchanged. Elevation data, for example, could be of real or integral in the DTM "files". The other is to be simplified or to be the form of grey level. The simplified form is coding. For example, coding for the soil types, 1 represents red soil, 1 represents sand soil, 3 represents alkaline soil, etc. The grey level form is to complete the transformations for the original data in terms of 256 grey levels. We can get 256 levels, for example, of the hight values in a concerning region, so the "lightest" grey level represents the largest hight value. Because of the constraints of computer storage capacity and storage exchanges with peripheral storages, one often uses the simplified auxiliary data. In the DTM image, for example, one grey value representing a elevation value needs one byte only for storage, so economizing the computer storages.

There are two kinds of employment modes for auxiliary data in the computer processing procedure. One is to be the fundamental data for the logical processing, the other is to be directly used as a auxiliary "band". DTM image, for example, looks like a frame of space remote sensing image, as the different grey level represents the different elevation value. So, it looks as the informations in a certain band of image and can reflect the relief and the running of mountain ridges. While we use it as a "band" for the classifications, there will be a kind of controlling effects of hights.

As to the fundamental data for the logical processing, there exist two kinds of applications. One is to be used for segmentations. For example, the extractions of spectral informations in a certain elevation belt, that depends upon DTM images. The other is to be used for dealing with the unclassified pixels in post processing. There are some "blanks" in the result images after computer aided classifications. Those are the pixels which computer can not "know" how to sort out. In this case, we can classify them in terms of some auxiliary data, for example, the grass growing on saline-alkali soil can be defined as Stipa-Artemisia.

Auxiliary data is the key factor in the computer aided classification procedure and can not be utilized without digital composition technologies. It means that the prerequisite for the use of auxiliary data is the correspondence in geometrical location of each pixel in a image with one of auxiliary data. For the pixel (i,j), what is the corresponding elevation value and what is the corresponding soil type? Those must be taken into account at first when digitizing the auxiliary informations. The digitalizations of soil type distribution map must correspond with remote sensing data window to be dealt with and one must implement the insertions and the densifications after the digitalizations of topographic maps. In short, one must make the Dm data consistant with the An data in geometry. Though it is not difficult to be taken into practice, some studies about that must be done in detail, the methods of insertion, for example. If we complete all of those processing, the computer aided classification precision will be much higher.

4. A successful result of application

In order to realize the grasslands classification principles during the computer classifications, we have completed some tests and then obtained the satisfactory results for the grasslands resources investigations of Bange county in the Northern Tibet.

This area is of 30,000 square kilometers. At the south of this county is Nien Qing Tang Gula mountain with the hight of more than 7,000 m. The

Bange town is located at the centre of this area. This county can be devided in five natural subareas, i.e. northwest subarea with more saline-alkali lands, southwest subarea with more wide valley plains and hills, northeast subarea with more lakes and flat grounds, and two zones in southeast where one is the high mountain zone (Nien Qing Tang Gula mountain), the other is the middle and lower mountain zone. The grasslands horizontal district distributions are very distinct in this area; from east to west, the rainfall decreases, and the grass group in the wide valleys changes from Stipa purpurea to Stipa purpurea-Carex moorcroftii, Kobresia-Stipa purpurea on the upper of mountains disappears gradually; from north to south, the terrain becomes higher and the grass group changes from Stipa purpurea to Kobresia-rank grass. Simultaneously, the grasslands district distributions are very clear, for example, Stipa purpurea-rank grass on the wide valley and flat ground, Stipa purpurea-Morina on the lakeside terrace, Stipa-Artemisia+Carex moorcroftii and Stipa purpurea-Carex moorcroftii round the saline-alkali lands, Achnatherum hookeri on the sand lands at dry river-bed, etc.

In order to make sure that there are enough spectral informations for the classifications, four frames of Landsat images in summer have been chosen which cover the whole county, dated July 1985, July 1983 and September 1984 (two frames to the last). Those four frames are depart from each other just at the neighbour of Bange town, thus one can consider those four frames as natural areas for which one will implement the classifications individually, not do that after putting together the four frames. Of those four frames, Lasa frame is at the southeast, and is divided in two zones in accordance of the principle on the geomorphologic subarea. Because this county is too large for the grasslands resources investigations, if classifying the whole data put together, the first, the calculations are too vast and it needs too much times to do the calculations again when getting the unsatisfactory results; the second, because of the date of image frames are not the same, and the contrast is different after putting them together, so the spectral reflections of same ground feature are not identical in those four image frames. In that case, the classifications will be failure; the third, according to the principle of visual interpretation, one must treat each zone in different strategies; the fourth, the geomorphologic subarea must be taken into account for the second level of classifications. Consequently, it is necessary and must be followed to implement the classifications in each zone.

As the hight difference is of more than 2,000 m, the grasslands vertical distributions are very distinct, for example, in the east area, the vertical distribution is Stipa purpurea--Stipa purpurea-Kobresia--Kobresia-Stipa purpurea. So we must introduce the DTM into classifications. After digitalization of topographic maps, transform the DTM data into grey level values for forming a "auxiliary band" in classifications.

In order to get the better results, we have chosen band 5 and band 7, not used all of the image data. Those two bands of data concern the informations of biomass index as follows;

$$Ib = \frac{B7 - B5}{B7 + B5} \qquad (5)$$

Because the spectrum analysis of various typical ground features show us that the difference of vegetation reflections between band 7 and band 5 is the greatest, and that lusher vegetations grow, greater the difference are. For this reason, those two bands are very worthy for the grasslands investigations.

The computer classifications are carried out in the ARIES-III image processing system by using the maximum likelihood method as the classifier. When we select the training areas for various grass groups in every natural zone, we must simultaniously determine the prior probabilities for each group according to their distribution in every natural zone. Those have been done in terms of the photos and the sampling records from field works, combining with the analysis about the characteristics of tones, colors, spots of different classes and the relationships of their locations. After trail classifications, we must adjust the training area and the prior probability of some class when necessary. It can be seen here that the principle of third level grasslands investigations has been taken into practice and the vegetation conditions involving the appearence characteristics, the growth situations and the ground environments have been considered when selecting and adjusting the training area and prior probabilities, though it is a little rough.

With the Gan-shu Grasslands Ecology Institute as the cooperation unit, and a lot of field surveying materials offered by that institute as the reference data, we have determined 18 classes for classification, i.e. kobresia-Stipa purpurea (We name it in short as K-S. The following in the paranthesis is the same meaning.), Stipa purpurea-Kobresia (S-K), lakes, Stipa purpurea (S), Stipa purpurea-rank grass (S-R), Stipa purpurea-Morina (S-M), Stipa purpurea-Carex stenophylla (S-C), uncovered rock lands, gravel lands, Kobresia Littledalei (K), saline-alkali lands (S-L), Kobresia-rank grass (K-R), Stipa purpurea-Carex moorcroftii (S-C), Blysmus sinocompressus (B), Stipa-Artemisiat Carex moorcroftii (S-A), Achnatherum hookeri (A), bush clump, and snow.

One must ensure, of course, the reasonableness of the grasslands classifications in the joining parts of the neighbouring zones, escaping from the absoluteness.

Those processing, including the use of DTM, have taken into computer classifications the scientific principles on the grasslands investigations, involving the geographical concepts, and the regularities of grasslands distributions, Here, we have seen the simplified realizations of grasslands classification theories and the combination of the identity and the diversity.

After the necessary post processings, one can put together the results of five natural zones. If there appear some unreasonable spots at the joining parts of the neighbouring zones, the modifications have to be done. We have found out, when doing it, that where the considerations for joining parts are thoughtful, the spots are joined up very well, and look very natural. However, some are the conjunctions of the spots which are of the similar types of grasslands.

When the modifications are finished the county boundary can be put on to get the last results. And then we can complete the statistics about the area of every class in whole county or in each countryside.

In order to assess the last results, the experts of Gan Shu Grasslands Ecology Institute have examined that with all of the record data (26 sampling points) in the field works and found that the rate of the correct points reaches to 96.2% and the uncorrect points are of the deviations between the similar classes, not the very large errors. For example, the grasslands which ought to belong to the class (S-K) have been classified as that belonging to the class (K-S), and nearby the check point, there really exists this kind of spot. So it is only the displacement and one can say that it is a small deviation. All of those show us that the computer aided classification results are correct for the qualitative assessment and the distributions from the point of view of the entirety.

Furthermore, the statistics have been done for the quantitative assessment, taking a countryside with the area of 600 square kilometers as the check field and comparing the area of spots drown out when doing visual interpretations with that of the corresponding spots resulted from computer aided classification. The correct rate or accuracy reaches to 95.2%. That indicates the correctness of computer aided classifications as viewed from area statistics for various classes.

On the other hand, the efficiency of computer classifications for the grasslands surveying is very evident. If implementing the visual interpretations, it needs five persons who must work on it in one year to set up the interpreting signs, draw out the grasslands spots, transfer the theme in mapping and calculate the areas for the grasslands investigations of 30,000 square kilometers. However, it need only two months to fulfil the computer processing. It resolves, consequently, the serious problem of the vast labours for the larger area of grasslands surveying. Now, this result is offered to the concerning planning and production department for serving the economic constructions.

5. Conclusion

The discusses for taking into practice the grasslands classification theories during computer processing in first three parts, and the effects of computer classifications for the grasslands investigations in Northern Tibet show us that the "road" becomes wide and wide, along which we are going for the resources surveying by using remote sensing image data.

We have not stopped at the dead end of the "way" that uses spectral informations only for the classifications. It is not the fantasy that the experts thinkings is employed in processing automations of remote sensing images, and one has entered into the initial utilizations, and could do more and more things. The important factor is that the scientists who are engaged in remote sensing digital image processing must integrate with the experts who work in the other fields which concern part of the problems of the former. When he is engaged in a certain job about applications of remote sensing, he must learn a little special knowledges on that, and understand the theories, the thinking ways and the approaches of those experts. Then, he must consider how to create the conditions to take into practice the expert thinking based upon the present conditions, doing that starting at the simplified experts' thinkings. Those above have been proved by the success of computer classifications by means of auxiliary data and the flexible computer processing technologies for 30,000 square kilometers of grass resources investigations in Bange county in accordance with grasslands classification theories and the requirements for three level surveying. The realizations of experts' thinkings in the field of remote sensing is completely possible and could reach to the satisfactory effects.

Referance:

1. "The manual on grasslands jobs", edited by the Grasslands Department of Gan Shu Agricultural University, 1979.

2. Shu Ning and Guan Zhe-Qun: "The studies on the computer classifications for the grasslands investigations in Northern Tibet by using Spacelab colour infra-red images", the paper of the Congress on Photogrammetry and Remote Sensing in Si'An, Dec, 1987.