Philosophical Principles of Data Discovery

Quan Wu[1,2(✉)], Min Liu[1,2], Juanying Sun[1,2], Weijie Jiao[1,2],
Shuanghua Tao[1,2], Xiaochen Li[1,2], Xue Han[1,2], and Lijuan Jia[1,2]

[1] Key Laboratory of Cultivated Land Use, Ministry of Agriculture,
Beijing, People's Republic of China
477781686@qq.com, 512773973@qq.com, 331631052@qq.com,
327352947@qq.com, 757659767@qq.com, 1351123026@qq.com,
346334593@qq.com, 448705149@qq.com
[2] Chinese Academy of Agricultural Engineering, No. 41 Maizidian Street,
Chaoyang District, Beijing 100125, China

Abstract. Data discovery is a basis in science and technology research because of nothing to do in case of no data. So, it is very important to find data. Where is the data? Data exists in the research objects which include material and consciousness in Philosophy. The world is material and motion is the fundamental attribute of matter. Therefore, things, as study objects, are always moving and changing. Data is description of the attributes of things which exist in a relatively static situation or absolute motion mode. The discovery of data is not only to observe the static state of the object, but also to study the state of motion of things. Sometimes it is necessary to let objects move. Time and space are the basic existence forms of things. Therefore, it is advisable to explore the things' attributes from two dimensions of time and space. Timing data is the description of the relative static state of objects on every time node in the process of moving along the time axis. According to the research needs or experience, the time nodes in the timing data are artificially set. Physical shape is the representation of things in the motion process along the space axis. Things may behave differently in different forms. The data describing the different performance can be called Situation Data. This paper introduces Qinghai Lake and a place in Three River Plain in China through presenting their timing data and situation data.

Keywords: Data discovery · Philosophy · Material · Thing · Timing data
Situation data · Time · Space · Motion · Time node · Virtual reality
RS

1 Introduction

Philosophy is the world view and methodology. Dialectical materialism reveals the general rules of nature world and human society. 'Movement is the fundamental way of matter existence' and 'time and space are the basic forms of matter existence' as the basic content are presented in Dialectical materialism. Therefore, it is a scientific method to observe things from the three dimensions of time, space and movement. There is no doubt that this method is deserved to recommend, because it can help people to broaden their horizons and often get unexpected harvest. Data discovery is a

basis in science and technology research because of nothing to do in case of no data [1–4]. So, it is very important to find data. Where is data? Data exists in the research objects including material and consciousness in Philosophy. How to find data from research objectives? According to philosophical viewpoints mentioned above, it is a good method to observe consciously the objectives from the three dimensions of time, space and movement.

Things, as study objects, are moving and changing. The movement is the expansion of things in time and space. The attribute of things exists in the process of movement in the form of space and time. The relative stillness in absolute motion provides possibilities and opportunities for researching objects and finding data. In order to discover data, it is necessary not only to study the properties of things under normal conditions, but also use the time and space to explore things for getting the attributes of things at the different time points and in the different space forms. The former can be called Timing Data (TD), and the latter can be called Situation Data (SD). TD is the description of the relative static state of objects on every time node in the process of moving along the time axis [5–7]. SD describes the attributes of objects at one time point in different performance forms in the process of moving along the space axis.

It is obvious that the objective things in nature can only have a kind of spatial expression at a time node. In other words, SD is single in nature. However, under the condition that the scale of observation is constant, subdividing the time node can obtain multiple SD because time is infinitely divisible. Of course, according to the requirements and the conditions of the existence of things, multiple SD can be also got through the virtual reality technology even without subdividing time [8, 9].

2 The Experiment

According to the epistemology of dialectical materialism, scientific knowledge comes from the two leaps. The first leap is from perceptual concrete form to rational abstract while the second leap is from rational abstract to rational concrete form. The viewpoint of observing things from the three dimensions of time, space and movement comes from productive practice, which completes the first leap of knowledge. The next thing to do is to apply the idea to practice, to test its correctness, and to realize the second leap.

2.1 The Purpose of the Experiment

This paper tries to introduce the method and process of data discovery with applying the viewpoint mentioned above by two cases. Meanwhile, these two cases are also used to test the correctness of the viewpoint.

2.2 The Objects and Content of the Experiment

The objects of this experiment are two things that exist in nature, which had to accept human being intervention. One of these is Qinghai Lake located in Qinghai province in northwest China. The other is a place in Three River Plain in Heilongjiang province in northeast China. The main content of this study is to investigate the present situation

and development trend of these two objects. Through the study, the method of data discovery will show.

2.3 The Method of the Experiment

The first question is how to begin with the test. According to the viewpoint of this paper, it is undoubtedly logical to observe the objects from three dimensions of time, space and movement for finding data. In other words, this is looking for SD and TD based on observing carefully the objects. Of course, the data format is different from the research object and the research content.

2.3.1 The First Data Discovery

Because the objects of this study are the natural things in the specific geographical location, it is advisable to choose RS image data as the basic data format for further deriving SD and TD. According to the experience and test requirement, Landsat image is the first choice of RS image for the test. Based on Landsat archived data covering the test objects the observation dates were determined. Obviously, taking into account researching development trend of the test objects, three observation dates were selected. So, the RS image on each time point forms a SD which is grid data form. Meanwhile, the RS images on three time points together form a TD which is also grid data form. The data situation is shown in Table 1. The SDs and TDs describing the spatial distributions of the objects are listed below in Figs. 1 and 2.

Table 1. The experimental RS data

Location	Sensor	Path/row	Dates	Band combination
Qinghai Lake	Landsat TM5	133/34 133/35	2006/09/20	RGB:453
			2009/08/11	
	Landsat OLI		2016/07/29	RGB:564
A place in Three River Plain	Landsat TM5	114/27	2006/08/30	RGB:453
			2010/08/25	
	Landsat OLI		2015/09/08	RGB:564
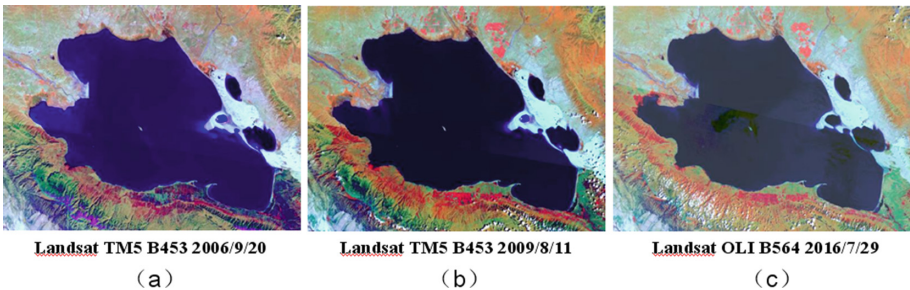
Landsat TM5 B453 2006/9/20   Landsat TM5 B453 2009/8/11   Landsat OLI B564 2016/7/29
          (a)   (b)   (c)

Fig. 1. The RS images of Qinghai Lake in Qinghai province in northwest China

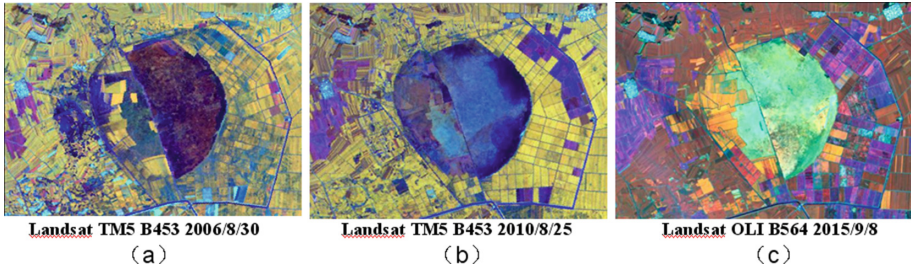Landsat TM5 B453 2006/8/30 (a)   Landsat TM5 B453 2010/8/25 (b)   Landsat OLI B564 2015/9/8 (c)

Fig. 2. The RS images of a place in Three River Plain in northeast China

The experimental data includes 6 remote sensing images covering Qinghai Lake and the place in Three River Plain in China for three years. The RS images of Qinghai Lake in 2006 and 2009 were made by Landsat TM5 while the image in 2016 was made by Landsat OLI. A complete Qinghai Lake RS image was made of two standard Landsat images which path/row number are 133/34 and 133/35. The band combination of each image is shown above in Table 1. The Fig. 1 shows the RS images of Qinghai Lake in Qinghai province in northwest China.

The RS images of the place in Three River Plain in 2006 and 2009 were made by Landsat TM5 while the image in 2016 was made by Landsat OLI. The band combination of each image is shown above in Table 1. The Fig. 2 shows the RS images of the place in Three River Plain in Heilongjiang province in northeast China.

The SD is (a), or (b), or (c) in Figs. 1 or 2. Meanwhile, the TDs are composed of (a), (b), and (c) in Figs. 1 or 2. The TDs and SDs in the form of RS images data individually describe the two test objects from the time axis and space axis. It is clear that, based on RS experience, some information, which is mainly qualitative, perceptual and concrete, can be gotten from the SDs and TDs in the form of grids.

2.3.2   The Second Data Discovery

It is the second data discovery to interpret the RS data. Interpretation is an important step in understanding remote sensing image [10, 11]. The target information can be obtained systematically through interpretation. This information can be generated in a different data format. The spatial distribution of Qinghai Lake in each year of three years can be accurately displayed with vector data form derived from interpretation. The each spatial distribution map which is (a), or (b), or (c) in Fig. 3, is a SD. The three maps form a TD consisted of (a), (b) and (c) in Fig. 3. The spatial distribution maps are shown below in Fig. 3.

The landuse situation of the place in Three River Plain each year of three years can be also accurately displayed with vector data form derived from interpretation. The each landuse map is a SD which is (a), or (b), or (c) in Fig. 4. The three landuse maps form a TD consisted of (a), (b) and (c) in Fig. 4. The landuse maps are shown below in Fig. 4.

The TDs or SDs exists in vector data format in the second discovery. The main features of these vector data is polygons. Since each polygon has a unique attribute and geographic location, the next step is to analyze the data through calculation based on
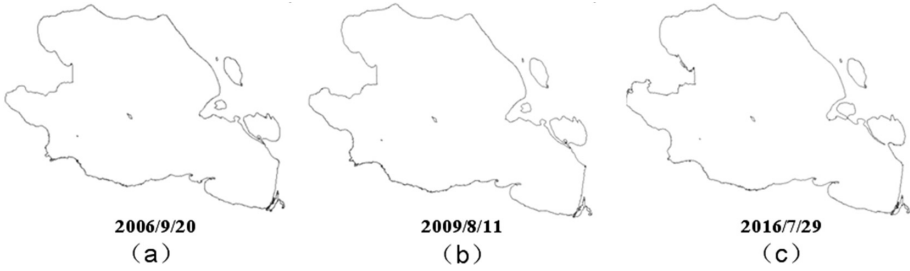
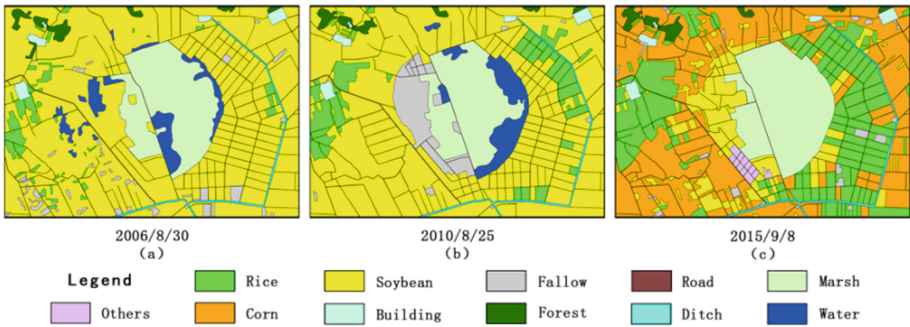**Fig. 3.** The spatial distribution maps of Qinghai Lake in three years

Fig. 4. The landuse maps of a place in Three River Plain in three years

GIS technology. It is a leap from perceptual concrete to rational abstraction that vector polygon graphs were generated from the RS images. By using the vector polygons, more information, which is mainly quantitative, can be achieved.

2.3.3   The Third Data Discovery

It is the third data discovery to analysis and synthesize the data secondly discovered. In the GIS platform, it is easy to calculate the area and number of each category polygon according to the polygon attribute classification by the year. The perimeters of polygons describing Qinghai Lake shape in three years were calculated. So, two tables shown below were generated from vector maps derived from the second data discovery. This process of generation of the two tables can be called the third data discovery which data format is a table form. The each column which is (a), or (b), or (c) in Tables 2 or 3, is a SD. The three columns form a TD consisted of (a), (b) and (c) in Tables 2 or 3.

In Table 2, the area and perimeter of Qinghai Lake in three years were precisely calculated respectively. Each SD accurately describes the status of Qinghai Lake's water surface, mountains in the lake and beaches in this year. The TD quantitatively shows the changing trend of Qinghai Lake from 2006 to 2016. An interesting discovery is that the water surface area is not reduced.

Table 2. The area and perimeter of Qinghai Lake in three years. Unit: hectare; meter

| | 2006/9/20 | | | 2009/08/11 | | | 2016/7/29 | | |
| | (a) | | | (b) | | | (c) | | |
	Number	Area	Perimeter	Number	Area	Perimeter	Number	Area	Perimeter
Hill	2	137.7	7352.5	2	137.0	7405.4	2	117.0	6455.6
Water	8	432847.3	569938.7	8	432323.0	568126.1	7	443955.8	591455.8
Beach	2	12.1	2599.8	2	114.4	7525.1	1	17.0	2660.5

Table 3. The landuse area of the place in Three River Plain in three years. Unit: hectare

| | 2006/8/30 | | 2010/8/25 | | 2015/9/8 | |
| | (a) | | (b) | | (c) | |
Landuse	Number	Area	Number	Area	Number	Area
Rice	47	517.3	48	1355.1	124	4002.7
Soybean	137	11538.1	149	10379.7	100	2020.8
Fallow	47	352.4	20	917.2	22	192.1
Forest	7	219.9	6	188.8	6	231.1
Building	5	169.9	5	169.9	6	171.6
Road	22	228.7	19	260.6	25	282.5
Water	12	620.4	4	705.4		
Ditch	1	166.3	1	166.3	1	166.3
Marsh	3	2219.9	2	1889.7	3	2569.1
Corn					115	6247.9
Others					12	149.4
Total	281	16032.9	254	16032.7	414	16033.5

In Table 3, the area and number of each type of landuse in the place of the Three River Plain were accurately calculated. Each SD accurately describes the landuse status of the region in this year. For example, in 2006 and 2010 there was no corn in the place, and in 2015 a lot of corn was planted. The TD quantitatively shows the landuse changing trend of the region from 2006 to 2015. The edge of the marsh had been abandoned to plant and the water surface area increased in 2010. But in 2015, the region had greatly increased the rice area and the land which was abandoned to plant was again planted soybean. Meanwhile, the water disappeared.

2.4 The Result of the Experiment

Based on SDs and TDs, one result of the experiment is that the water surface area of Qinghai Lake had not only decreased but increased in passed ten years. Another result of the experiment is that the place in Three River Plain was being used to adjust the landuse structure, and the marsh in the middle of the place has a tendency to disappear. Through the above test results, the relevant departments can adjust policies and take measures to control the future development direction of the two regions.

3   Conclusions

It has important significance to apply consciously the viewpoints of time and space of things in philosophy, because it is possible and easy to observe things from the multidimensional space where things exist for finding and constructing SDs and from the development history where things exist in time for finding and constructing TDs. The SD describes the status of things at a certain point in time. Because the space of things is multidimensional, the SD can be varied. Multiple SDs of the same type constitutes the TD that can show the basic law and direction of the development of things. The more SDs is, that means the time sequence is longer, the more accurate the rules and directions which TDs reveal are. Whether it is research or production applications, it is possible to understand things and then grasp the development direction of things with the SD and the TD.

References

1. Mariscal, G., Marban, O., Fernandez, C.: A survey of data mining and knowledge discovery process models and methodologies. Knowl. Eng. Rev. **25**(2), 137–166 (2010)
2. Fu, T.C.: A review on time series data mining. Eng. Appl. Artif. Intell. **24**, 164–181 (2011)
3. Ignatiadis, N., Kiaus, B., Zaugg, J.B., Huber, W.: Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. Nat. Methods **13**(7), 577–581 (2016)
4. Grandell, T.: Visual information discovery. Inf. Serv. Use **36**, 57–63 (2016)
5. Ares, J., Lara, J.A., Lizcano, D., Suarez, S.: A soft computing framework for classifying time series based on fuzzy sets of events. Inf. Sci. **330**, 125–144 (2016)
6. Toshniwal, D.: Feature extraction from time series data. J. Comput. Methods Sci. Eng. **9**, 99–110 (2009)
7. Rodpongpun, S., Niennattrakul, V., Ratanamahatana, C.N.: Selective subsequence time series clustering. Knowl.-Based Syst. **35**, 361–368 (2012)
8. Coxon, M., Kelly, N., Page, S.: Individual differences in virtual reality: are spatial presence and spatial ability linked? Virtual Reality **20**, 203–212 (2016)
9. Budziszewski, P., Grabowski, A., Milanowicz, M., Jankowski, J.: Workstation for people with disabilities: an example of a virtual reality approach. Int. J. Occup. Saf. Ergon. **22**(3), 367–373 (2016)
10. Simms, D.M., Waine, T.W., Taylor, J.C., Brewer, T.R.: Image segmentation for improved consistency in image-interpretation of opium poppy. Int. J. Remote Sens. **37**(6), 1243–1256 (2016)
11. Saraf, A.K., Mishra, P., Mitra, S., Sarma, B., Mukhopadhyay, D.K.: Remote sensing and GIS technologies for improvements in geological structures interpretation and mapping. Int. J. Remote Sens. **23**(13), 2527–2536 (2010)