

Automatic Generalization of Residential Areas Based on “Paradigm” Theory and Big Data

Tianlin DUO^{a,1} and Peng ZHANG^b

^a *Henan University of Technology, 50 Lianhua Street, High Tech Development Zone, Zhengzhou, Henan, China*

^b *School of Mechanical and Electrical Engineering, Henan University of Technology, 50 Lianhua Street, High Tech Development Zone, Zhengzhou, Henan, China*

Abstract. “Paradigm” theory is an important ideological and practical tool for scientific research. The research means and methods of Geographic Information Science follow the laws of four paradigms. Automatic cartographic generalization is not only the key link of map making, but also a recognized difficult and hot issue. Based on large-scale map data and deep learning technology, an automatic cartographic generalization problem-solving model is proposed in this paper. According to the key and difficult problems faced by residential area selection and simplification, residential area selection models and simplification models based on big data and deep learning are constructed respectively, which provides new ideas and schemes to solve the key and difficult problems of residential area selection and simplification.

Keywords. Date intensive paradigm, large date, geographic information science, automatic cartographic generalization, deep learning, feature extraction

1. Introduction

Cartographic generalization is the key link in map making and the main method to solve the contradiction between the area of the ground and the map. No matter map making, database deriving, multi-scale expression of spatial data and updating of spatial data, cartographic generalization should be carried out whenever scale transformation is involved. At present, the level of automatic cartographic generalization has been greatly improved. However, there are still many problems worth studying. Based on the understanding of the development law of Geographic Information Science, the inspiration of the research trend of big data and data intensive paradigm, the analysis of machine learning and deep learning technology, and the carding of the difficulties and hot issues in cartographic generalization research, this paper puts forward the idea of solving the difficulties involved in automatic generalization based on large-scale map data and deep learning technology.

¹ Corresponding Author, Tianlin DUO, Henan University of Technology, 50 Lianhua Street, High Tech Development Zone, Zhengzhou, Henan, China; Email: dtl@haut.edu.cn.

Project Source: Doctoral research project of Henan University of Technology (2020BS034).

2. The Scientific Significance of “Paradigm” Theory and Jim Grey’s Four Scientific Paradigms

Both “paradigm theory” and “scientific theory” are concepts in the field of philosophy of science. With the continuous development of science and technology, the concept of “scientific theory” gradually shows its limitations. Kuhn put forward the “paradigm theory” which is more theoretical generality and accurately discusses the trend of scientific development. Generally speaking, scientific theory is a pure knowledge system or a pure logical structure without social and historical factors; Kant, Popper and others emphasized the unmanned, truth and objectivity of scientific theory; There is no scientific theory that can explain all the problems in this research field, and no scientific theory can cover all the problems in this research field. The objective phenomena are complex and the theoretical coverage is limited. It can be seen that scientific theory ignores the objective existence of human beings and the relationship between human beings and scientific development. According to the view of social history, only scientific theory which is accepted by a certain group of people can be called scientific theory. What has nothing to do with people is not science.

Based on the above background, Kuhn put forward “paradigm theory” and sorted out the mode of scientific development. The structure of scientific revolution is Kuhn’s earliest work on paradigm theory, which insists on the view of historical materialism and social practice on the progress of science.

Paradigm refers to the set of beliefs, values and technologies shared by members of a community. It refers to the theoretical basis and practical norms on which conventional science operates. It is the world outlook and behavior pattern that researchers engaged in a certain science abide by [1-2].

Kuhn’s view on paradigm is of great significance to scientific development and scientific research activities. Once the paradigm theory is put forward, it has been unanimously recognized by the academic circles. This paper summarizes the essence of science, sorts out its context and trend, so as to effectively guide scientific and technological workers to carry out scientific research [3].

Jim grey further put forward four paradigms theory, summarized the research paradigms of each stage of scientific development, and put forward the general time period, research means and methods of each stage. Today’s and future research in various fields will be characterized by data intensive paradigm. That is, based on spatiotemporal big data, with Internet, Internet of things and cloud computing as technical means, through big data analysis and mining, we can find new patterns, new knowledge and new laws that can not be found by previous research methods [4].

3. The Significance of Big Data for Scientific Research

Data intensive science paradigm is based on big data. Mining the value of big data has become the consensus of all countries, and developed countries have launched their own big data strategies. In the past two years, China has successively issued the “action plan for promoting big data development” and “big data industry development plan (2016-2020)”. Big data has achieved unprecedented success in agriculture, commerce, finance, transportation, tourism, medical treatment, social management, military and other fields. The significance of big data for scientific research lies in:

(1) Big data promotes the transformation of scientific research methods, is the

supplement and development of traditional scientific research methods, and provides new research ways and means for scientific research under the background of big data.

(2) Due to the rapid improvement of the ability of computer to acquire, store and process data, people gradually find the laws that are difficult to find from a large number of data, and the data becomes the basis of decision-making.

(3) In the development process of various fields, a large amount of data has been accumulated, most of which have not been effectively managed, analyzed and utilized. With the improvement of information processing technology, a new scientific research field data intensive science will emerge.

Academician Wang Jiayao pointed out that in today's geo information science, "Internet plus" big data and "GIS" evolution and development report, the current geo information science is facing global challenges in the era of big data. The arrival of the big data era is also an important factor in promoting the development of Geographic Information Science.

4. Problems in Automatic Cartographic Generalization

At present, spatiotemporal big data analysis and mining has become the most competitive theory and technology in the field of geographic information science. Cartographic generalization is a difficult and hot issue in this field. With the unremitting efforts of many experts and scholars, cartographic generalization has experienced a series of evolution from objectification to quantification, modeling, algorithmic, and collaborative and systematic based on Algorithms and models. The level of intelligence and automation of cartographic generalization has been improved to a certain extent. Due to the complexity of cartographic generalization and the constraints of research methods, there is still a long way to go to the requirements of real automation and intelligence [5-7].

Cartographic generalization is carried out under a series of rules. In many cases, the process of cartographic generalization depends on experts' experience and thinking in images, which is difficult to model and algorithmize. Now there are also some automatic generalization systems based on expert knowledge base, but the generalization rules are formalized one by one, which is difficult to exhaust and the workload is huge. The emerging machine learning based methods are in the initial stage, there have been some achievements, but also the problem of feature extraction.

5. Automatic Cartographic Generalization Problem Solving Mode Supported by Big Data

The research of data intensive paradigm and big data has become an inevitable trend in various fields. Can the existing large-scale map data be used for deep mining and analysis in the field of cartographic generalization to solve the problems that are difficult to be solved by conventional methods. First, we should analyze the problems involved in automatic synthesis [8-10].

5.1. Key Issues Involved in Automatic Synthesis

In the problem of factor selection, there are many scientific methods for quota selection, such as load capacity, root mean square, linear regression model, etc. The difficulty of automatic synthesis lies in the modeling and quantification of expert experience in structural selection. Simplification, generalization and displacement involve feature extraction and graph matching [11-12].

With the efforts of many research teams for many years, we have achieved a series of breakthrough results, such as the application of genetic algorithm, mathematical morphology, neural network, graph theory, case learning, decision tree and other methods. However, there are still many problems worthy of study in this field. The key is feature extraction, especially automatic feature extraction to reduce manual annotation [13-14].

5.2. Problem Solving Mode of Automatic Cartographic Generalization Supported by Big Data

There are four basic forms of cartographic generalization: selection, simplification, generalization and displacement. Taking the residential area of a certain area as an example, the paper takes selection and simplification as the experimental object, and uses the excellent feature learning ability of multi hidden layer artificial neural network to extract the image thinking law of experts which is difficult to quantify [15].

When selecting point residential areas, the input parameters are grade weight, position importance, shape importance, density eigenvalue and conflict factor; When selecting the area residential area, the input parameters are the grade of residential area, the importance of population, the importance of location, the degree of dispersion and the area of residential area as shown in figure 1.

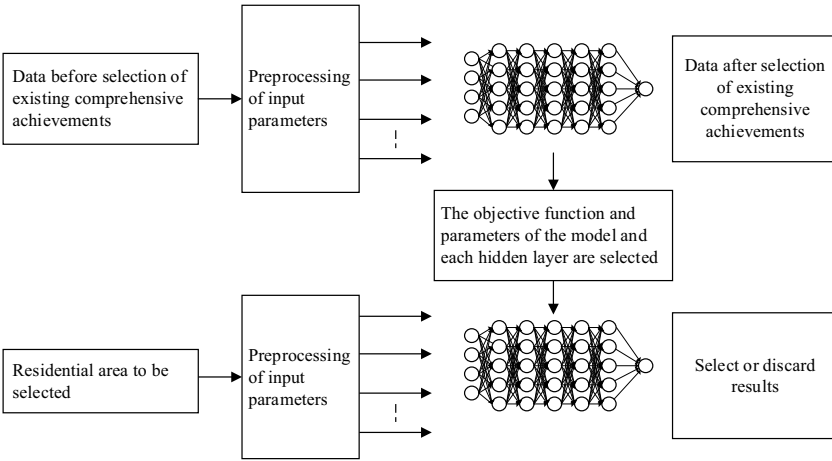


Figure 1. Residential area selection model based on big data and deep learning.

When the residential land is simplified, the excellent feature learning ability of the multi hidden layer artificial neural network is used to detect the residential land similar to the residential land to be simplified. The results of different experts in the same

residential land are different. Using the idea of statistical language model in natural language processing for reference, the simplified result with high probability is used as the simplified result of the residential land to be simplified, as shown in figure 2.

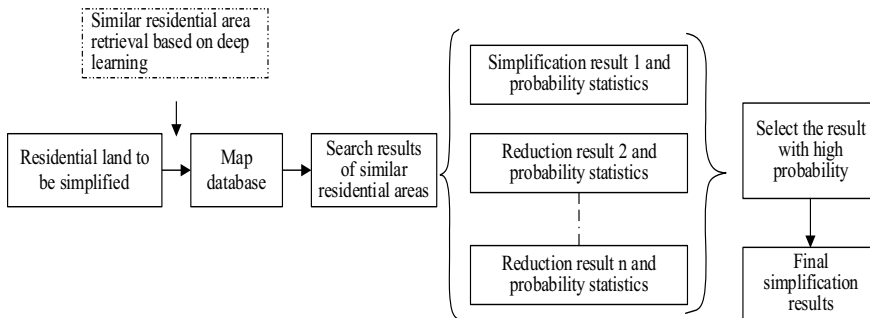


Figure 2. Residential land simplification model based on big data and deep learning.

6. Conclusion

The transformation of scientific research methods brought about by the era of big data is a supplement and development to traditional research methods. The "paradigm" of data intensive scientific research has become a consensus in the scientific community. Deep learning is good at exploring complex structures in high-dimensional data, and has shown good prospects in the fields of image recognition, speech recognition and so on. Effective use of a large number of existing cartographic generalization achievements and the advantages of automatic feature extraction by deep learning provide a new solution for the intelligent research of automatic generalization. Based on the in-depth analysis of the key and difficult problems existing in residential area selection and simplification, residential area selection models and simplification models based on big data and deep learning are designed respectively. In terms of specific methods, many technical problems need to be further studied. With the in-depth research, the research ideas based on big data and in-depth learning must show their advantages.

References

- [1] Chen M.comparative study of “paradigm” and “scientific theory” [D].Wuhan University of science and technology, 2011
- [2] R.K. Merton. Sociology of science [M]. Beijing: Commercial Press, 2003:20-25
- [3] Kuhn, Jin WL and Hu XH. The structure of scientific revolution [M].Beijing: Peking University Press, 2005: 15-19
- [4] Did Li Xingmin Kuhn. use the term “incommensurable” for the first time in the philosophy of science[J] Science, technology and dialectics, 1994, (04)
- [5] Wang JY.Progress and Prospect of cartography and geographic information engineering [M].Edited by China Society of Surveying and mapping. Blue book on the development of Surveying and mapping in China (volume 2008). Beijing: Surveying and Mapping Press, 2008:13-19
- [6] Liu ML,Zhou L.rethinking the definition of Geographic Information Science[J]. Geography and geographic information science, 2004.20 (3): 1-5
- [7] Wang JY, Cheng Y, Wu MG, Sun QH. Evolution and development of g7eographic information system[J].Journal of Surveying and Mapping Science and technology, 2008 (4): 4-9

- [8] Zhang YZ. Four kinds of thinking brought by big data. Learning times, 2015-1-26 (F8ourth Edition)
- [9] Wu F,Qian HZ,Deng HY,et al.Intelligent processing of spatial information for automatic map generalization[M].Beijing: Science Press, 2008:1-5
- [10] Hu HM,Qian HZ,He HW,et al.Automatic selection of residential areas based on analytic hierarchy process[J].Journal of Surveying and mapping, 2016.45 (6): 740-746
- [11] Huang Y et al. Optimal Lipschitz maps on one-holed tori and the Thurston metric theory of Teichmüller space[J]. Geometriae Dedicata, 2021, : 1-24.
- [12] Li CM et al. An area merging method in map generalization considering typical characteristics of structured geographic objects[J]. Cartography and Geographic Information Science, 2021, 48(3) : 210-224.
- [13] Yang JT et al. Identifying Complex Junctions in a Road Network[J]. ISPRS International Journal of Geo-Information, 2020, 10(1) : 4-4.
- [14] Ma D et al. PolySimp: A Tool for Polygon Simplification Based on the Underlying Scaling Hierarchy[J]. ISPRS International Journal of Geo-Information, 2020, 9(10) : 594-594.
- [15] Shen YL and Ai TH. A Raster-Based Methodology to Detect Cross-Scale Changes in Water Body Representations Caused by Map Generalization[J]. Sensors, 2020, 20(14)