Knowledge-based Generalization on Land-use Data

GAO Wenxiu[1] GONG Jianya[1] LI Zhilin[2]

[1.] National Laboratory of Surveying, Mapping and Remote Sensing, Wuhan University

129, Luoyu RD. Wuhan, China, 400079

[2.] **Department of Land Surveying and Geo-informatics**

The Hong Kong Polytechnic University

Abstract

Land-use data is one important form of categorical data. The generalization of land-use data has been the subject of many literatures, but most of them focused on data model and algorithms. The purpose of the paper is to illustrate knowledge involved in the generalization of land-use data. In the test, knowledge is classified based on categorical data. Then the specific knowledge is collected for generalization of land-use data and is represented as rules with the Production-rule method. And a knowledge-based framework of the land-use data generalization is proposed. A case study shows the results of the experiments built based on the knowledge-based framework of land-use data generalization.

Key words: Land-use data generalization, generalization knowledge, thematic knowledge, general knowledge, Production rule, generalization rule

1. Introduction

Generalization of land-use data works for abstracting the main and important information from the original database or maps to build a new database or make small-scale maps to represent the more general characteristics of land-use in a wider range.

Knowledge-based generalization has been discussed in many literatures. Muller and Mouwes(1990) stated that the knowledge is required in the generalization process of topographic map. Nickerson(1991) revealed a set of rules mainly based on criteria of minimal sizes for the selected map objects based on the manual generalization process of the *Canadian National Topographical Series*. Armstrong(1991) proposed three knowledge types considered in performing cartographic generalization, which are geometrical, structural and procedural knowledge. Heisser(1995) described five knowledge groups on the basis of topographic map generalization: knowledge of geometry and graphics, semantic knowledge, procedural knowledge, structural knowledge and interdisciplinary expert knowledge. Kilpelainen(2000) identified four main rules for the generalization of Finnish topographic maps: geometric, topological, context-related and culture-related rules.

Generalization of land-use data has been the subject of many literatures. Monmonier(1983) promoted a raster-mode approach to realize merging, splitting, eliminating and partitioning area object for land-use and land cover map generalization. Goffredo (1995) studied low-level and high-level generalization using both raster and vector domain procedures for land-cover automatic generalization systems. Olli Jaakkola (1997) implemented generalization of land-cover data with a raster modeling and converted different datasets together. Liu (2002) took land-use data generalization as an example to study the model generalization of categorical database in GIS. He also discussed some constraints in categorical database generalization.

The purpose of the paper is to study the knowledge involved in generalization of land-use data. First, a knowledge classification is illustrated based on generalization of land-use data. Then the methods of knowledge representation are simply discussed. In the study of the paper, generalization knowledge is represented as rules with the Production-Rule methods of artificial intelligence. A case study shows some generalized results based on the knowledge. Although developed to transform the land-use data from larger-scale to smaller-scale, the knowledge classification and generalization rules discussed here are applicable to other categorical data, such as vegetation data and soil data.

2. Knowledge classification

This study explores the generalization knowledge for land-use data from two aspects:

General Knowledge

- Graphic/geometrical knowledge:
- Topological knowledge: Keep right topological relationship after generalization including polygon-polygon relationship, polygon-line relationship, polygon-point relationship, line-line relationship, line-point relationship, and etc.;
- Generalizing operations knowledge: Knowledge about generalizing operations. right orders of different generalizing operations, and etc.;
- Knowledge of data management platform: database, data model, data structure, data handling, and etc.

Thematic Knowledge

- Nature-based knowledge: Difference laws of geographic terrain, distribution states, relationship between different geographic feature classes, and etc.
- Culture-based knowledge: Relationship between mankind and earth, related policies or rules, social and economic benefits, and etc.
- Application-based knowledge: Research region, research theme, map scale, purpose, and etc.

Knowledge representation is one of the important components when designing a knowledge-base generalization system. There are various knowledge representation techniques such as Production Rules, Logic, Semantic Network, Frame, State Space, Conceptual Dependency, Script, and etc. The test here selects Production Rules to represent knowledge for land-use data generalization. The syntax of the Production Rules:

IF < conditions / antecedent > : THEN < operation / consequent >

This structure is composed of the left-side conditions or antecedent which is a logical combination of propositions about the database and right-side operation or consequent which contains a collection of actions or states(Shea 1991, Armstrong 1991). And this structure is termed a production rule. The antecedent of a rule states a condition or aspect of the problem which must be present in a application, while the consequent specifies an available actions to solve the problem. When executing a production system, a rule is triggered if its antecedent is matched, but the triggered rule is merely examined to be an available rule not to be executed right now. When all available rules have been examined, the most appropriate rule is chosen and fired to execute the action of its right-side part.

3. Knowledge for land-use data generalization

(1) Geometric Knowledge

The aim of geometrical knowledge is to maintain clear graphic representation of map objects, which generally employs the minimal thresholds of graphical size. The graphical thresholds include minimal area, minimal length, minimal density, and minimal interval distance between two objects, minimal curve on line objects or boundary of area objects. For the land-use data, the thresholds are changeable for different land-use types. Table 1 shows the detailed information of the thresholds in the study, which came from the related documents of manual generalization of land-use map. And the thresholds are changeable with different scales.

Table 1 The geometrical threshold

Type of Threshold	Type of land-use	Threshold (Variable Name)
Area of land parcel	Resident & Construction Area	4 mm^2 (MiniArea1)
	Farmland & Garden Plot	6 mm^2 (MiniArea2)
	Woodland, Grassland, Unused-land	15 mm^2 (MiniArea3)
	Lake, Reservoir, Pool Plot	1 mm^2 (MiniArea4)
Length of line object	River, Canal	11 mm^2 (MiniLength1)
	Rural Road	15 mm^2 (MiniLength2)
Interval Distance between objects	Land Parcels	1 mm^2 (MiniDistance1)
	River, Canal, Rural Road	3 mm^2 (MiniDistance2)
	General Highway	5 mm^2 (MiniDistance3)
Minimal curve	Land Parcels	2mm * 2mm (MiniCurve1)
	Lake, Reservoir, Pool Plot	0.8mm*1.0mm(MiniCurve2)

(2) Importance ranks of land-use types

Importance ranks of land-use types represent their positions in economic society of studied regions and the important ranks impact the orders and methods of generalization actions. There is various knowledge to determine the importance ranks of land-use types such as nature-based knowledge, culture-based knowledge, application-based knowledge.

(3) Rules of selection:

Selection means to select related feature types and features for a specific application. Rules of selection should consider various factors such as the importance ranks of land-use types, map scale, purpose of application, data modeling of GIS system, geometric knowledge and other thematic knowledge. The following rules show some examples for selection.

For polygon objects:

IF *polygon objects are full of researched region*

THEN *select all the polygon objects*

For line objects, different rules specify different types of line objects.

railway, highway, general Highway: generally select all railways and highway

IF *an object is railway or highway or general highway*

THEN *select it*

Rural Road: select those which is longer than the minimal length(MiniLength2)

IF *an object is rural road* ***AND*** *its length > MiniLength2*

THEN *select it*

River, Canal: select those which is longer than the minimal length(MiniLength1)

IF *an object is river or canal* ***AND*** *its length > MiniLength1*

THEN *select it*

(4) Rules of attribute transformation:

Attribute transformation is a main step for database generalization or model generalization. Its task is to transfer related data from the original database to a derived database. The database generalization is regarded as the content transformation of a spatial database from a high resolution to a lower resolution representation (Molenaar 1996). For land-use data, the main task of the attribute transformation is to convert the land-use type from a detailed level to a generalized level based on a classification hierarchy. Figure 1 shows a classification hierarchy of land-use types. From the first level to the third level, the information becomes more and more generalized. And we call the land-use types in the second and third levels the parent of the relative land types in the first level.

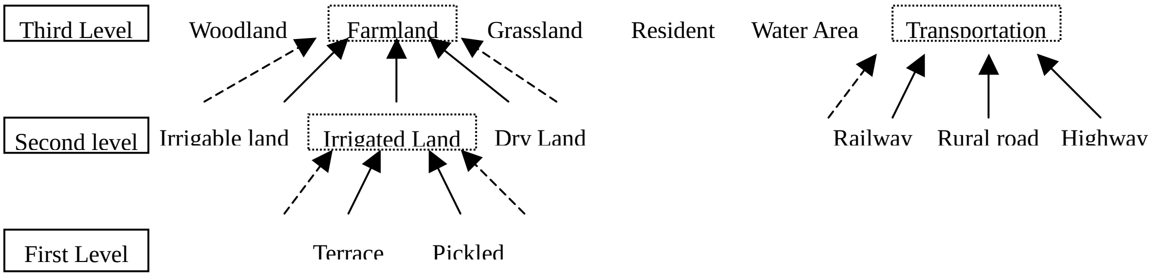

Figure 1 Classification hierarchy of land-use type (Part of it)

The attribute transformation is governed by some constraints. The first one is the transformation is one-way conversion from the first level to second level or third level. The second one is to conform to the semantic consistency between the new type and the old type of a land-use object. That is to say, “Terrace” land should be converted into “Irrigated Land” of the second level or “Farm land” of the third level, rather than “Dry Land” or “Grassland”. The third one is that the important land-use types can be expressed as a different level from the minor land-use types. The transformation emphasized the important land-use types.

IF *the new class level is greater than old class level*
AND *the new class is the parent of the old class*
THEN *Transform land-use type of the object*

(5) Rules of merge:

After the attribute transformation, the adjoining land parcels belonging to the same land-use type should be merged into one land parcel. For the land-use data, there is another constraint for merging operation: although two adjoining parcels have the same land-use type, the two parcels can not be merged into one because they belong to two different administrative regions. This constraint will avoid the mistakes on area statistics of administrative regions.

IF *two land parcels are adjoining*
AND *belong to the same land-use type and same administrative region*
THEN *Merge the two parcels into one*

(6) Rules of operations on small land parcels:

In order to operate the small land parcel, the premise is to identify which land parcels are smaller than the thresholds. Table1 shows the minimal area thresholds for each kind of land-use type respectively. The following rule shows an example to identify which land object is a small one.

IF *an object is used as farmland*
AND *Its area < MiniArea2*
THEN *the object is a small farmland parcels*

For small land parcels, different important ranks have different operation strategies. And for different spatial status and land-use types, the methods are also changeable. The following section will discuss how to cope with small lakes, pool plots and farmlands with adjoining relationship and containing relationship.

The land-use professional knowledge does not allow to aggregate several small lakes and pool plots into one or more ones.

IF *an object is a lake or pool* ***THEN*** *BeAllowedAggregation = FALSE*

If a small land parcel adjoins other land parcels, identify if there are any adjoining land parcel which has the same parent land-use type as that of the researched land parcel. If there are some, collapse the researched land parcel and average its area to those land parcels. If only one adjoining parcel has the same parent land-use type, just aggregate the two parcels. The purpose of this strategy is to retain the total area of this kind of land-use unchanged.

IF *$object_i$ is adjoining $object_j$* ***AND*** *they belongs to the same parent land-use type*
AND *BeAllowedAggregation = TRUE*
THEN *Aggregate them into one parcel*

If no any adjoining land parcel has the same parent land-use type as that of the studied land parcel, different strategies could be adopted to different important rank. For the important land parcels, enlargement is the first choice. If there is no enough space around it for enlargement, check if it is possible to displace the land parcel first. Then enlarge it to the minimal area size again. If the displacement might cause invalid topological relationship, the strategy could not be implemented. The last choice for important land parcel is to collapse it and average its area to its adjoining land parcels.

If a small land parcel is contained completely in another land parcel, we call it the island-parcel and the other land parcel the background-parcel. If some small island-parcels in a background-parcel belong to the same land-use type, aggregate them into one or more large island-parcels if they are close enough to each other. If these island-parcels have the same parent land-use type, aggregation is also a good choice.

IF *$object_{i,}$ ……, $object_i$ are contained completely by a background-parcel*
AND *$object_{i,}$ ……, $object_i$ belong to the same land-use type*
THEN *Aggregate objecti, ……, objecti into one object*

The single important island-parcel should be enlarged if the space is enough.

IF *$object_i$ is important* ***AND*** *there is enough space for enlargement*
THEN *Enlarge the object*

If enlargement causes a new conflict, the single important island-parcel could be amalgamated with the background-parcel.

IF *$object_i$ is important* ***AND*** *there is no enough space for enlargement*
THEN *Amalgamate $object_i$ with its background-parcel*

For the lake and pool plot, aggregation and deletion are not suitable, while enlargement, amalgamation and collapse could be the good order to cope with them.

4. A case study

The case study demonstrates the importance of knowledge-based generalization. The basic platform for the test is Arc/info. The land-use data tested here is collected from rural regions of Guangdong Province of China and is organized with area objects (land parcels) and line objects (rivers, roads etc.), both of which are put in different coverage in ARC/INFO. The purpose of the case study is to transform 1:5000-scale land-use data to 1:10000-scale and 1:50000-scale land-use data. The 1:5000-scale land-use data presents the land-use information of the rural area and the smallest administrative level is countryside. The 1:10000-scale presents the land information of town-level administrative regions. The 1:50000-scale presents the land information of county-level administrative regions. (In China, the ranks of administrative region are country, province, city, county, town, countryside.)

Figure 2 shows a part of the original map. The numbers marked on the land parcels are the codes of land-use type. Figure 3 shows the generalized result without rules of merge after attribute transformation, that is, the adjoining polygons with the same land-use types (such as 53 or 55) are merged although they belong to the different administrative regions. Figure 4 shows the generalized result with the rules of merge. The adjoining parcels with same land-use type were not merged because they belong to the different administrative regions. The results of figure 3 will cause wrong area statistics of each administrative region.

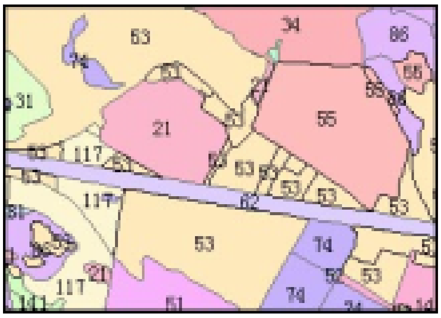

Figure 2 Original sample data

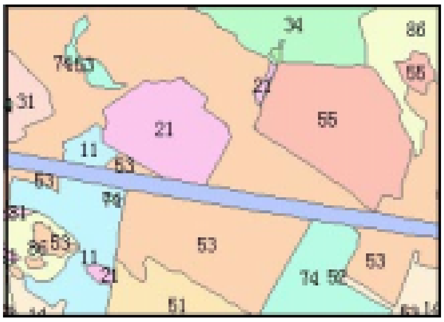

Figure 3 Results without Merge Rules

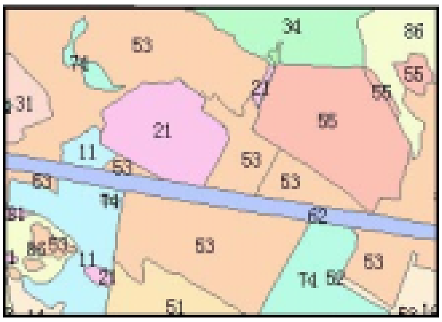

Figure 4 Results with Merge Rules

Figure 5 presents that the four 1:5000 land-use maps are zoomed into 1:50000 without any generalization. Figure 6 presents the result after generalization based on the above rules. Table 2 shows the area statistics values before and after the generalization.

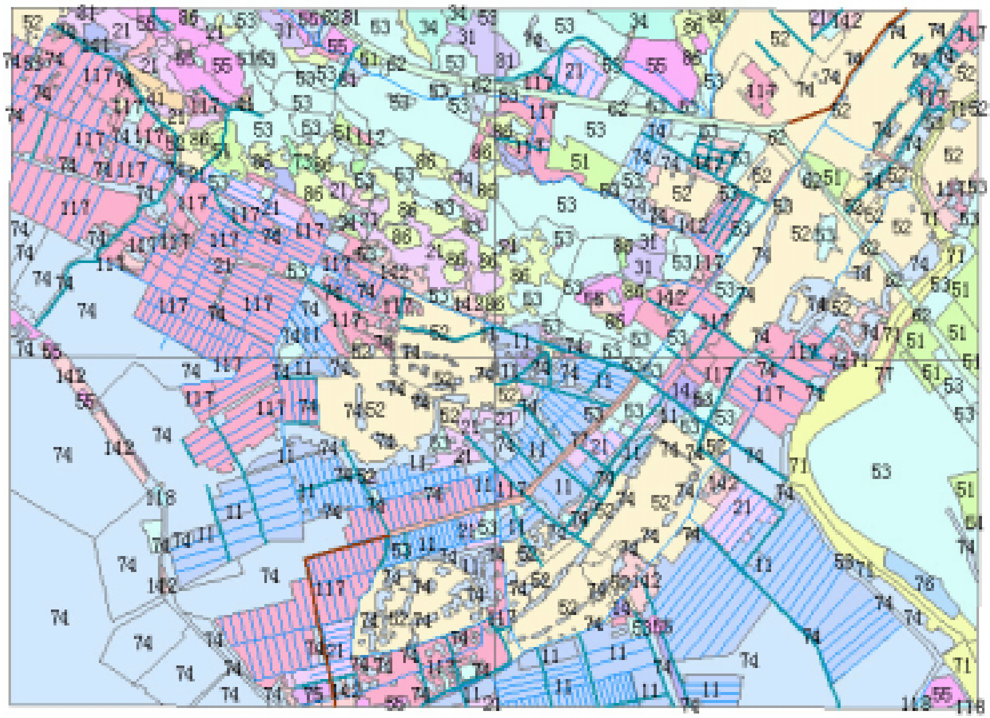

Figure 5 Four 1:5000 land-use map is zoomed into the size of 1:50000 without ant generalization

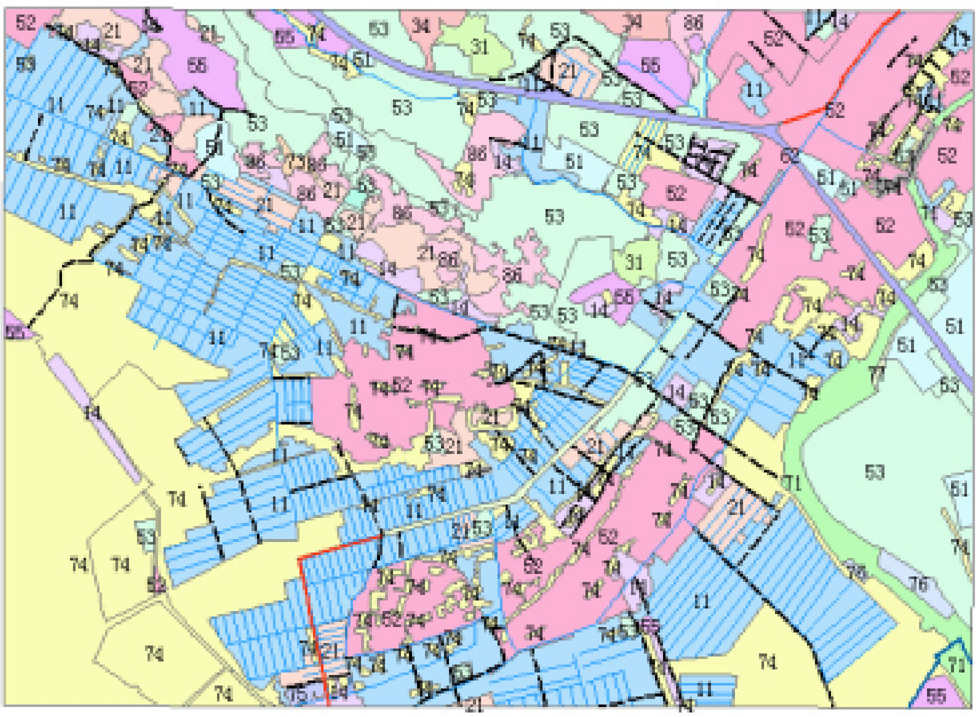

Figure 6. 1:50000 land-use map after generalization based on the above rules

Table 2 the area statistics values before and after generalization

Total area of researched field	Land-use Type (code)	Area before generalization(m^2)	Area after generalization(m^2)	Area Rate before generalization	Area Rate after generalization
29482308.14	Farmland(1)	8630560.18	8760673.42	29.3%	29.7%
	Garden Plot(2)	1130087.65	1144565.48	3.8%	3.9%

	Woodland(3)	422747.09	349291.09	1.4%	1.2%
	Grassland(4)	48804.30	0	0.17%	0
	Resident(5)	10258777.54	10545012.41	34.8%	35.8%
	Transportation(6)	272822.31	279104.04	0.93%	0.95%
	Water Area(7)	7838651.54	7719598.9	26.6%	26.2%
	Unused Land(8)	879857.53	684063.61	2.98%	2.32%

Conclusion drawed from Table 2:

(1) The area rates of different land-use types have almost kept an abalance before and after the generalization. The biggest change rate is 1%.

(2) After the generalization, the area rate of different land-use types still presents the importance ranks of the land-use types metioned above. (Resident always accounts for the biggest area in most of studied field in China).

(3) There are 1027 land parcles before the generlaizaiton, only 306 land parcels are in the map after the generalization. The rate of the area reduction is 70%.

5. Remarks for the future

Land-use data is a complex categorical data. It not only involves point, line, polygon features, but also it strongly relates to the cadastral information management. Generally, in practice a land information database is required to satisfy with many applications such as land-use plan, cadastral information management, land-use change etc. And these applications all require multi-scale or multi-detail land information, so a reasonable generalization system of land information database is the key problem. Generalization knowledge is the basic stone and indispensable components for a generalization system. The paper focuses on analysis and collection of knowledge for land-use data generalization and employs the knowledge in practical experiments. The results proved the importance of thematic knowledge. But the contents of thematic knowledge for land-use data generalization are changeable for different practical regions. So the preferential method should be to build an open system so that users can add their knowledge when they execute data generalization and the system can extend the knowledge base with the additional knowledge. As a good generalization system, identifying automatically spatial knowledge according to geographic features in database is another important function, which is the only valid way to save time and improve the precise of generalized results.

Reference

1. Armstrong, M. 1991, Knowledge Classification and Organization, in Buttenfield, B.P. and McMaster, R.b.(Eds) Map Generalization: Making Rules for Knowledge Representation, pp.87-102, London: Longman
2. Goffredo S. 1995, Knoweldge-based system for automatic generalization of satellite-derived thematic maps. Proceedings of the 17th International Cartographic Conference, Vol.1, pp. 108-117
3. Heisser, M., Vickus G. and Schoppmeyer J., 1995, Rule-orientated definition of the small area 'Selection' and 'combination' steps of the generalization procedure, GIS And Generalization: Methodology and Practice, in Muller J. C., Lagrange J. P. and Weibel R.(Eds.), pp. 135-147
4. Kipelainen T. 2000, Knowledge Acquisition for Generalization Rules, Cartography and Geographic Information Science, Vol.27, No.1, pp.41-50

5. Molenaar Martien, 1996, Mult-Scale Approach of Geodata, Int. Archives of Photogrammetry and Remote Sensing, Vol, XXXI, B3, pp. 542-554
6. Muller, J.C. and Mouwes P.J., 1990, Knowledge acquisition and representation for rule based map generalization: An example from the Netherlands, GIS/LIS Proceedings 90, Anaheim, California, Vol1, pp.58-67
7. Nickerson B.G. , 1991, Knowledge Engineering for Generalization, in Buttenfield, B.P. and McMaster, R.b.(Eds) Map Generalization: Making Rules for Knowledge Representation, London: Longman, pp. 40-55
8. Olli Jaakkola, Multi-Scale land cover databases by automatic generalization, *Proceedingsof 18th ICA/ACI International Cartographic Conference*, Stockholm Sweden, June 1997, Vol. 2 pp.709-714

Information of the first author:
GAO Wenxiu
National Laboratory of Surveying, Mapping and Remote Sensing
Wuhan University
129, Luoyu RD. Wuhan, P.R.China, 400079
TEL: 0086-27-87881292
FAX: 0086-27-87643969
Email: wenxiugao@hotmail.com
wxgao@mail.liesmars.wtusm.edu.cn

Research activities on generalization:
Apr. 2001- Feb. 2002, Researched on the project of vegetation map generalization in **Department of Land Surveying and Geo-informatics** of The Hong Kong Polytechnic University with Dr. PUN Lilian and Dr. LI Zhilin.
The title of her Ph.D. thesis is “Research on Knowledge-based Thematic Data Generalization in GIS” and she graduated in Dec. 2002.