MODELING UNCERTAINTY LAND-USE DATA IN VIETNAM

Tran Nhu Trung

Vietnam General Department of Land Administration (GDLA)

Email: TranITC@hotmail.com

ABSTRACT

Due to in proper land management, land information are captured and processed with uncertainty. This is difficult to avoid in Vietnam, where land management is administrated through four levels, namely state, province, district and commune level. Thereby various resolutions of land use data and generalization is required. In order to overcome these inconsistencies, uncertainty in land use data has to be modeled and applied for different resolutions.

This paper aims at working out a modeled which presents uncertainty in land use data at any resolution, whereby land use data can be stored in database with corresponding uncertainty and resolution. The model is then created on the analysis of various uncertainty sources in condition of Vietnam. An Entity Diagram (ER) of data model is modified to fit with uncertain data. Every step, in which data is adapted and modified, is investigated; so that a model of uncertainty propagation can be created and finally can be tested, using real data captured from the Nhankhang commune in Vietnam.

1. INTRODUCTION

The discrepancies between the real world and mapping result of GIS often exit. This mismatches is always a part of GIS and is known as uncertainty topic. The uncertainty of land use might be the issue of land use classification, or the issue of ontology between land use and land cover (Fisher 2000). Another reason causing uncertain land use data is the mix-use of land by land-user. (Trung 2001) shows that habit of living and farming at the same area creates uncertainty in boundary.

Furthermore, such uncertainty data is used in multi-resolution of Vietnam land administration. And to support the land use generalization task (from high to lesser resolution), (Trung 2001) proposed four solutions to apply. In fact, those solutions open the way to accept mix-use in land then to work with error when applying automatic generalization procedures. However, (Trung 2001) did not present the ability to handle uncertainty from raw data in use at each level of generalization.

Purpose of this paper is to develop a formal model for uncertainty of land-use data in Vietnam condition, and to present uncertainty at any resolution of database. In the following parts, section 2 introduces several cases of uncertainty of land-use data in Vietnam. Section 3 proposes data model to capture uncertainty. Section 4 discusses how uncertainty changes at each step of generalization. A case study to validate proposed is described in section 5. As usual, conclusion and suggestion for further research are presented in section 6.

2. UNCERTAINTY LAND-USE DATA IN VIETNAM

In Vietnam, three cases below normally causes to uncertainty in land use data.

Case 1: The overlap between land-use classes can cause uncertainty in both thematic and spatial parts (Molenaar 1998). For example, a land parcel used to grow eucalyptus tree can be put into forestry class or annual use class, see figure 1.

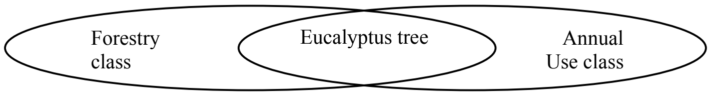

Figure 1: Class overlap leads to uncertainty

Case 2: Land use unit (LuU) is movable. Because of when doing land allocation, only purpose of using land and their extent (area) and general boundary of LuU are registered. Therefore land-user can change or move LuU in side general boundary. For example, farmer moves their residential land-use area to the near new road, figure 2.

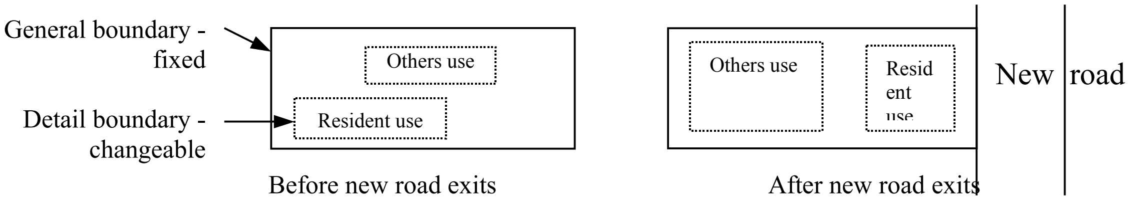

Figure 2: LuU moves when knowing new road exits

Case 3: Too many land use purposes exit at the same area can makes difficult to define where is boundary of each land use. Three purposes of using land like Resident, Garden, and Pond normally go together in rural of Vietnam. The bound of Resident, Garden and Pond is not crisp but the general bound is clear and registered in database, see figure 3 and (Trung 2001).

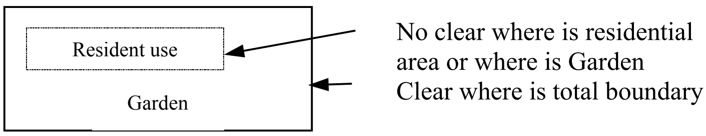

Figure 3: Uncertainty boundary for Resident and Garden

To summary, the uncertainty of case 1 is difficult to recognize in our database for reason we have no information about this situation. For case 2 and 3, the fact is that database only can capture geo-object with a certain boundary. That means LuU in database is LuU with general boundary. Therefore, uncertainty land use data means LuU contains many land use purposes at same time. In short, LuU has relationship M:N with land use class (LuC).

3 DATA MODEL FOR UNCERTAINTY LAND-USE DATA IN ENTITY DIAGRAM

The relationship between LuC and LuU is M:N. This relationship is normalized into two relationships 1:M and M:1. A new entity of relationship LuU/LuC is created - uncertainty entity, figure 6. Uncertainty table captures information of the ability one LuU belongs to one LuC. This ability is presented by function membership M, (Molenaar 1998).

Certainty that GO is belong to class A = M(GO, class A) [0-1] (1)

Suppose uncertainty land-use classification of a GO is a set (S), which contains of all membership class (C) that this GO might belong to.

Scertain (GO) = { M(GO, class 1),... , M(GO, Classn) (2)

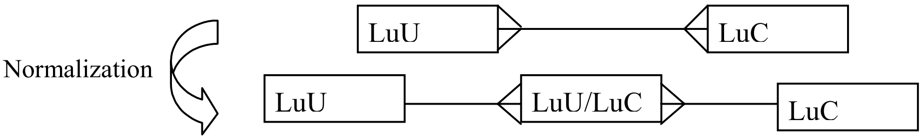

Figure 4: Relationship of LuU and LuC in uncertainty case

4 UNCERTAINTY LAND-USE DATA IN MULTI-RESOLUTION ENVIRONMENT

These are four types of generalization: Class driven generalization, structure, geometric and functional generalization, (Molenaar 1996, 1998). In domain of land-use data, class driven is most suitable to be used, (Trung 2001). Class driven generalization contains two spatial actions: (1) Geo Object (GO) changes its class to supper class, (2)It is aggregation of all GO (with the same class and are neighbor) to form a new GO. Uncertainty propagates differently at each above steps. Those steps are analysis following:

4.1 Uncertainty propagation in Class generalization

Class generalization means class of GO change its class to supper class but GO boundary still keeps constant. To formal the relations of $S_{certain}$ (GO) before and after generalization, two cases below are investigated.

Case 1: For classes that are children class of K (or K is those supper class).

Suppose A is an action. And Case $_1$, Case $_2$, …, Case $_n$ are cases that A can happen. Ability for each case is correspondently to: A_1, A_2, …, A_n. If K is case that unions all those cases.

K = Case $_1$+ Case $_2$ + ... + Case $_n$. Then the ability action A happens in case K will be A_1+A_2+..+A_n]. This is the same rule for the case that all classes under super class. Therefore:

$M(GO, Class\ K) = \sum M(GO, class\ K_i)$ (3) In which class K_i is children class of K.

Case 2: For classes is not children class of K (or K is not their supper class).
In this case we found that when GO follows class generalization to class K, the ability GO belong to class K does not depend on the ability GO belongs to other class - outside class K.

To handle two cases above, set S is divided into two sub-sets. One contains all membership class that under class target K and other is the rest (outside K).

$$S_{certain} (GO)= S_{certain} (GO)_{under\ K} + S_{certain} (GO)_{outside\ k} \quad (4)$$

If F_{gk} is function for doing class generalization with class K and $S_{certain} (GO)_{AF}$ is uncertainty data after doing. Then:

$$S_{certain} (GO)_{AF}=F_{gk} (S_{certain} (GO))= F_{gk} (S_{certain} (GO)_{under\ K} + S_{certain} (GO)_{outside\ k}) \quad (5)$$

Applying result of case 1 & case 2 above, the final formula is:

$$S_{certain}(GO)_{AF} = \sum M(GO, class\ K_i) + S_{certain}(GO)_{outside\ k} \quad (6)$$

This form shows that after class generalization, only uncertainty part under class driven will change, but the others outside will resistant.

4.2 Uncertainty propagation in aggregation step

When aggregating GO, new object will contain all uncertainty from originated object. (Cheng 2001) shows uncertainty of new object is directly proportional to area of member object and their uncertainty. That means uncertainty after merging can be calculated as:

$M(GO, class\ K) = \sum M(GO_i, class\ K) x Area\ of\ GO_i / Area\ of\ GO$ (7)

Let PA_i = Area of GO_i / $\sum$ Area of GO_i (8)

Then $M(GO, class\ K) = \sum [M(GO_i, class\ K) \times PA_i]$ (9)

This can be applied to the set of uncertainty data in (2). If $S_{certain}(GO)_{AF}$ is set of uncertainty that changes after doing aggregation.

$S_{certain}(GO)_{AF} = \{ \sum [M(GO_i, class\ 1) \times PA_i], \ldots, \sum [M(GO_i, class_n) \times PA_i$ (10)

Base on (6) and (11) final formula presenting the relations of uncertainty before and after generalization is deducted.

$S_{certain}(GO = \sum [\sum M(GO_i, class\ K_i) \times PA_i] + \sum [S_{certain}(GO_i)_{outside\ k} \times PA_i]$ (11)

5 EXPERIMENTS

Purpose of this experiment is to validate the ability of proposal model in keeping uncertainty in land-use data at any resolution. Study area is a part of Nhankhang commune, Hanam/Vietnam. In this area, cadastral surveying was taken and LuU was mapped. The status of uncertainty land-use data is presented on table 1 and figure 5.

- Capture uncertainty land-use classification: Table 2 contains thematic of LuU. Table 3 is created for uncertainty with the same structure as section 3. Primary key is combination of LuID and LuC, column *uncertainty* contains the ability that LuU belongs to a class. This *uncertainty* is calculated following equation (9).
- Process class driven generalization: Five classes are required to be generalized: Agriculture, Forestry, Residential, Special use and Unused. Applying formula (6), table 3 (uncertainty table) will change to table 4. After doing aggregation, new objects are captured in table 5. New table of uncertainty is created - table 6 - calculated following formula (9). Result of class driven generalization is shown on table 5,6 – thematic, uncertainty information and graphic as figure 7.

Result of experiment shows: (1) uncertainty on land-use data can be presented in database in very details of classes as table 2 or less detail as table 6; (2) equation 11 allows to present the relations of uncertainty between high and low resolution in land-use class.

Table 1: Legend & statistics data

Legend	Uncertainty on class of use	Number of land-use units
	Residential, Garden, and Aquatic use	73
	Residential, Garden	242
	Garden, Aquatic use	1
	Residential, Aquatic use	1
	No-mixed use (no uncertainty)	162

Figure 5: Nhankhang cadastral map

Figure 6: Uncertainty situation

Table 2: Land-use units

LuID	Area
11	884
12	470
17	445
18	550
19	376
...	

Table 3: Uncertainty table

LuID	LuC	Uncertainty
11	Residential	200/894 = 0.23
11	Garden	219/894 = 0.25
11	Aquatic	474/894 = 0.52
12	Crops	470/470 = 1.00
17	Residential	180/445 = 0.40
17	Garden	265/445 = 0.60
18	Residential	180/550 = 0.33
18	Garden	370/550 = 0.67
...		

Table 4: Uncertainty change in class generalization

LuID	LuC	Uncertainty
11	Residential	0.23
11	Agricultural	0.77 (=0.25+0.52)
12	Agricultural	1
17	Residential	0.4
17	Agricultural	0.6
18	Residential	0.33
18	Agricultural	0.67

Table 5: Land-use units

LuID (new)	Area
1	8,829.0
2	3,162.2
3	2,713.8
4	912.8
5	5,505.5
6	1,273.0

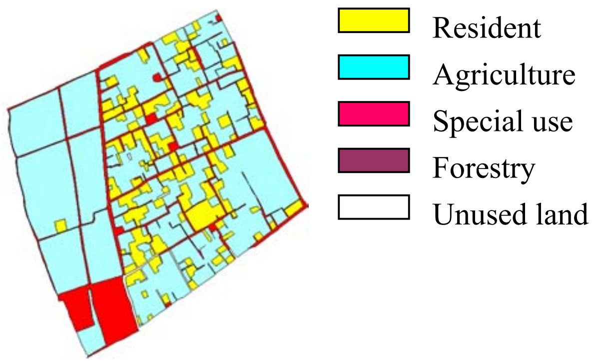

Figure 7: Result of class driven generalization

Table 6: Uncertainty after doing class driven generalization

LuID	LuC	Uncertainty
1	Residential	0.30
1	Agricultural	0.70
2	Agricultural	1
3	Residential	1
3	Agricultural	0.90
5	Residential	0.22
5	Agricultural	0.78
...		

6. DISCUSSION AND FURTHER RESEARCH

To conclude, the proposed model on section 3 and 4 have ability to capture uncertainty of Vietnamese land-use data. Uncertainty was modeled and kept in table 2,3. Table 2 and table 3 present land-use data in high resolution; table 5 and table 7 present for lesser resolution.

Interesting point in this result gained by compared uncertainty table 6 and table 3 (high & low resolution). This shows that the gap between maximum and minimum class in table 6 is much more than in table 3. For instance, the gap of maximum and minimum on table 3 (before generalization) is normally 0.4-0.6, but for table 6 (after generalization) is 0.3-0.7. It means that when change from high to lower resolutions, land use data becomes more certain. In the other words, generalization increases the certainty of data.

However, this result opens another question that what types of indicators should be used to measure the uncertainty at each level of generalization? Furthermore, this paper just limited on one type of database generalization – class driven. How this formal can be applied for other types of generalization: such as geometric, structure and functional driven? Above questions are worth for future research.

REFERENCES

Fisher, P., Comber, L., Wadsworth, R. 2001. Land-use and Land Cove: Contradiction or Complement.

Molenaar, M. 1998. An introduction to the theory of spatial object modeling for GIS. Taylor & Francis, London, 1998.

Cheng, T. 2001. Indicator of uncertainty in Multi-scale database. SDH 2001 Beijing.

Trung, T.N. 2001. Solutions to apply class driven generalization for land-use planning and land registration in Vietnam situation. Proceeding of Seminar on "GIS, Land management for sustainable development" in Hanoi. 16-17/2001. Organized by FIG & VGCR. Chaired by Prf. Dr. Dang Hung Vo and Prf. Tommy Österberg.