Estimating semantic uncertainty in land cover classifications - Methods and implications for data generalization

OLA AHLQVIST

DEPARTMENT OF PHYSICAL GEOGRAPHY

STOCKHOLM UNIVERSITY

S-10691 STOCKHOLM

SWEDEN

FAX: +46 8 16 48 18

E-MAIL: OLA @NATGEO.SU.SE

Scale change by spatial aggregation is technically a straightforward process. Harder though is to produce information from the aggregation that is consistent with the analysis purpose at the target scale. Despite the large interest in data aggregation and generalization, very few, if any, studies have investigated digital generalization effects with a quantitative, location specific approach. There is today no documented knowledge if there is a semantic variation in the use of a specific concept at different scales. That is, whether the human mind makes different interpretations of an area if we apply different scales for our investigation.

In this paper, the use of a location specific assessment of aggregation results provide an estimation of scaling and generalization effects on the semantic accuracy of categorical data sets. To enable this the reference dataset is derived using exactly the same manual interpretation technique and exactly the same input information as the for the collection of lower resolution data. This design is chosen to enable an evaluation of whether there is a semantic effect involved in aggregation of geographic information. These differences will here be estimated as differences in semantic accuracy.

Consequently, this work illustrates methods to detect semantic differences in a scale change from fine resolution, pixel based landscape data generalized into a coarse resolution data set. Due to the inevitable interpretation inconsistency, this study also implements methods to represent classification ambiguity using rough set theory. The anlysis uses methods of rough classification and error matrix extensions, and thus the potential for these novel methods to perform generalization and accuracy assessments is also illustrated.

The experiment demonstrates how to translate categorical uncertainty into rough classifications. The analysis further illustrates the possibility with rough classifications to do assessments of the correspondence between two datasets using semantically uncertain data in one or both of the assessed data sets. The evaluation of scale dependency in the use of certain landscape concepts such as 'agricultural district' and 'forest district' provides no clear evidence for a scale dependent use of these concepts. However, the findings still raise important questions for future quality assessments of digitally aggregated data. Used methods are therefore suggested as a way to estimate semantic accuracy. It is finally argued that more extensive testing is required of a possible scale dependency of commonly used land cover mapping concepts. The issues of

semantic accuracy assessments as well as contemporary techniques to produce digital land cover classifications are also disussed.