Behavior-based aggregation of land categories for temporal change analysis

Safia Zakaria Aldwaik a, Jeffrey A. Onsted b, Robert Gilmore Pontius Jr c,∗

a Ramallah Municipality, 4 Issa Ziadeh Street, Ramallah 600, Palestine
b Department of Earth & Environment, Florida International University, 11200 SW 8th Street, Miami, FL 33199, USA
c School of Geography, Clark University, 950 Main Street, Worcester, MA 01610-1477, USA

ARTICLE INFO

Article history:
Received 29 May 2014
Accepted 12 September 2014

Keywords:
Aggregation
Category
Change
Land
Map
Transition

ABSTRACT

Comparison between two time points of the same categorical variable for the same study extent can reveal changes among categories over time, such as transitions among land categories. If many categories exist, then analysis can be difficult to interpret. Category aggregation is the procedure that combines two or more categories to create a single broader category. Aggregation can simplify interpretation, and can also influence the sizes and types of changes. Some classifications have an a priori hierarchy to facilitate aggregation, but an a priori aggregation might make researchers blind to important category dynamics. We created an algorithm to aggregate categories in a sequence of steps based on the categories’ behaviors in terms of gross losses and gross gains. The behavior-based algorithm aggregates net gaining categories with net gaining categories and aggregates net losing categories with net losing categories, but never aggregates a net gaining category with a net losing category. The behavior-based algorithm at each step in the sequence maintains net change and maximizes swap change. We present a case study where data from 2001 and 2006 for 64 land categories indicate change on 17% of the study extent. The behavior-based algorithm produces a set of 10 categories that maintains nearly the original amount of change. In contrast, an a priori aggregation produces 10 categories while reducing the change to 9%. We offer a free computer program to perform the behavior-based aggregation.

© 2014 Elsevier B.V. All rights reserved.

Introduction

Purpose

Categorical scale concerns the level of detail of a set of categories. Categorical scale is also known as thematic scale, which describes the type of information that a map shows. The choice of categorical scale is one of the central challenges for geographers in mapping land cover, land use, vegetation type, soil type and other categorical variables. Comparison of maps that show a categorical variable at two time points for the same study extent can reveal change during the time interval. However, if maps have a large number of categories, then analysis can become difficult to interpret. Category aggregation is the process of merging detailed categories to create a smaller number of broader categories. Category aggregation can simplify interpretation, but can also reduce the amount of apparent temporal change, depending on which categories are aggregated. If category aggregation is performed strategically, then aggregation can play an important role in data mining, because strategic category aggregation can help to reveal information that might otherwise be lost by other types of aggregations. This article presents a new algorithm to perform a sequence of categorical aggregations in a manner that gives insights concerning categorical change over time. Our algorithm is based mainly on the temporal behavior of each category in terms of its gross gain and gross loss.

Literature

Briassoulis (2000) describes various classification systems that are popular for land change science. Many of these systems have a hierarchical structure, whereby detailed categories are grouped under a smaller number of conceptually broader categories. Anderson et al. (1976) established a popular system where the detailed categories can be aggregated to a coarser level in a hierarchy based on similarity of land use. For example, if we...
have various Agricultural subcategories classified according to the detailed Anderson level II system, we can move to the coarser Anderson level I system by aggregating all Agricultural subcategories into one broader Agricultural category. Aggregation can have important implications for subsequent analyses. For example, if the subcategories transition with each other, then aggregation will eliminate those transitions. Also, if one subcategory is responsible for change while other subcategories persist, then aggregation will lose information concerning which subcategory is responsible for the change. Conway (2009) demonstrated that various category aggregations can influence the calibration and validation of a land change simulation model. Aggregation can also influence the results of pattern metrics (Ahlqvist and Shortridge, 2010; Buyantuyev and Wu, 2007; Buyantuyev et al., 2010).

Table 1 is a glossary of terms that the literature and the remainder of our manuscript uses. We define these terms in a broad manner so that they have applications beyond land change science.

Pontius and Malizia (2004) investigated the influence of category aggregation on measurement of change over time. They derived five mathematical principles that dictate the effects of aggregation on the net change and the swap change, which are two components that sum to the total change. Net change is the component of change that derives from a difference between two time points in the number of pixels of each category. Swap change is the component of change where a category appears to reallocate. Reallocation occurs when a category experiences both gross gain in some pixels and gross loss in other pixels during the time interval. The five principles rely on characterizing each category according to the category’s net change, thus categories are labeled as net gainer or net losers. A net gainer is a category for which its gross gain is greater than its gross loss. A net loser is a category for which its gross loss is greater than its gross gain. Principle 1 dictates that aggregation cannot increase the total change. If the aggregation is either a net loser with another net loser or a net gainer with another net gainer, then Principle 2 dictates that the net change is maintained and Principle 3 dictates that the swap change decreases by the sum of the transitions between the aggregated categories. If a net loser is combined with a net gainer, then Principle 4 dictates that the net change decreases and Principle 5 dictates that the swap change can decrease, increase, or be maintained. These principles derive from the fundamentals of set theory.

This article introduces a new algorithm to perform a sequence of category aggregations that are based exclusively on the behavior of the categories. Our goal is to produce a new type of aggregation based on information that existing aggregation systems ignore. The mathematical foundation of our approach is the framework concerning net change and swap change, which is also known respectively as quantity difference and allocation difference (Pontius and Milliones, 2011). Specifically, our algorithm uses the second and third principles of Pontius and Malizia (2004) to maintain net change and to maximize swap change at each step in the aggregation sequence. Thus if individual detailed categories are involved in large transitions, then those categories tend to be maintained as unaggregated until latter steps in the aggregation sequence.

Case study in Redland, Florida, USA

We illustrate the characteristics and advantages of our algorithm with a case study from Redland, Florida, USA. Redland is part of the Florida Coastal Everglades (FCE) site of the United States National Science Foundation’s Long Term Ecological Research network. Redland is 20 miles southwest of Miami and between two national parks. FCE researchers are interested in the rapid suburbanization in some locations and conservation in other locations. Various government agencies are involved in conservation through zoning and acquisition of farmland via eminent domain. Researchers at FCE are examining the impact that zoning has on the calibration of a cellular automaton urban growth model (Onsted and Roy Chowdhury, 2014). Researchers have also been interviewing Redland residents regarding land use with respect to pesticides, fertilizers and water (Harris et al., 2012). We chose Redland as a study area because its data contain 64 categories, thus it illustrates well the need for aggregation. The Redland land categories are organized in an a priori hierarchy based on similarity of use, therefore this hierarchy can serve as an initial framework for use-based aggregation. However, the use-based aggregation can make scientists blind to important signals of land change that might otherwise be apparent with a different combination of aggregated categories. The remainder of this paper compares the use-based aggregation versus behavior-based aggregations that derive from our algorithm.

Methods

Data

We use raster maps from two time points to illustrate our algorithm. Each raster has 270 columns and 233 rows of pixels at the 200-m resolution, where each pixel is classified as exactly one category. Fig. 1a shows maps of 64 categories at year 2001 and year 2006. The rows of Fig. 2 show how the data have an a priori hierarchical structure that groups the categories conceptually into ten broader categories based on use. Fig. 1b shows these ten use-based categories. We compare the use-based aggregation to the behavior-based aggregations that our algorithm produces. The columns of Fig. 2 show the 10-category aggregation that our behavior-based algorithm produces. We present Fig. 2 now so that the reader can understand our goal. Fig. 1c shows these ten categories that our algorithm produces.

<p>| Table 1  |
|------------------------|------------------------|</p>
<table>
<thead>
<tr>
<th>Term</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contingency table</td>
<td>Square table where the number of rows and the number of columns equals the number of categories. The entry in row i and column j of the table gives the size of the study extent that is category i at the initial time and category j at the final time.</td>
</tr>
<tr>
<td>Transition</td>
<td>A particular off-diagonal entry in the contingency table.</td>
</tr>
<tr>
<td>Total change</td>
<td>Sum of all off-diagonal entries in the contingency table. Total change can be expressed as the sum of two components called net and swap.</td>
</tr>
<tr>
<td>Net change</td>
<td>Component of change that is attributable to differences in the quantity of each category between the initial time and the final time.</td>
</tr>
<tr>
<td>Swap change</td>
<td>Component of change that is attributable to gross gain of a category in some locations and gross loss of the same category in other locations during the time interval.</td>
</tr>
<tr>
<td>Exclusive loser</td>
<td>Category that has positive gross loss and zero gross gain.</td>
</tr>
<tr>
<td>Exclusive zero</td>
<td>Category that has zero gross loss and zero gross gain.</td>
</tr>
<tr>
<td>Exclusive gainer</td>
<td>Category that has zero gross loss and positive gross gain.</td>
</tr>
<tr>
<td>Swapping loser</td>
<td>Category for which its positive gross loss is greater than its positive gross gain.</td>
</tr>
<tr>
<td>Swapping zero</td>
<td>Category for which its positive gross loss equals its positive gross gain.</td>
</tr>
<tr>
<td>Swapping gainer</td>
<td>Category for which its positive gross loss is less than its positive gross gain.</td>
</tr>
</tbody>
</table>
The input to our algorithm is a square contingency table, for which the rows list the categories at the initial time point, the columns list the same categories at the subsequent time point, and the order of the categories is identical in both lists. The number of rows equals the number of columns in the table, even when a category's size is zero at one of the time points. Each entry in the table gives the number of pixels that belong to the entry's row category at the initial time and the entry's column category at the subsequent time. Each entry on the diagonal of the table shows persistence for a category, and each entry off the diagonal shows transition from the row category to the column category.

**Algorithm**

Table 2 gives the mathematical notation that our algorithm's equations use. Eq. (1) defines $L_k$ as the gross loss for category $k$ while Eq. (2) defines $G_k$ as the gross gain for category $k$. Eq. (3) defines $T_{ij}$ as the size of the sum of the two transitions from $i$ to $j$ and from $j$ to $i$. Eq. (4) defines $D$ as the total change among the categories between the initial time point and the subsequent time point. Eq. (5) defines $N$ as the net component of change, which derives from the difference between time points in the size of each type of category. The summation in the numerator of Eq. (5) double counts the total change because each pixel of change is loss of one category and gain of another category, therefore the 2 in the denominator of Eq. (5) neutralizes the double counting. Eq. (6) defines $S$ as the swap component of change, based on the fact that the overall change is the sum of the net component and the swap component.

$$L_k = \left( \sum_{j=1}^{J} C_{kj} \right) - C_{kk}$$  \hfill (1)

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I$</td>
<td>Number of categories.</td>
</tr>
<tr>
<td>$i$</td>
<td>Index for a category at the initial time point.</td>
</tr>
<tr>
<td>$j$</td>
<td>Index for a category at the subsequent time point.</td>
</tr>
<tr>
<td>$C_{ij}$</td>
<td>Size of study extent that is category $i$ at initial time point and category $j$ at subsequent time point.</td>
</tr>
<tr>
<td>$L_k$</td>
<td>Gross loss of category $k$ during time interval.</td>
</tr>
<tr>
<td>$G_k$</td>
<td>Gross gain of category $k$ during time interval.</td>
</tr>
<tr>
<td>$T_{ij}$</td>
<td>Sum of two transitions between categories $i$ and $j$.</td>
</tr>
<tr>
<td>$D$</td>
<td>Overall change during time interval.</td>
</tr>
<tr>
<td>$N$</td>
<td>Net component of overall change.</td>
</tr>
<tr>
<td>$S$</td>
<td>Shift component of overall change.</td>
</tr>
</tbody>
</table>
The behavior-based algorithm aggregates individual categories so that net change is constant, because the algorithm never aggregates a net losing category with a net gaining category. Each step in the behavior-based algorithm aggregates two categories so that theswap change is maintained as much as possible. The algorithm aggregates the category pair for which $T_{ij}$ is the minimum among pairs of categories that have the same net direction.

![Table 3](https://example.com/table3.png)

Fig. 3 shows how we use Eqs. (1) and (2) to assign each category $k$ in the original classification to one of six behaviors. Exclusive Loser means $L_k > 0$ and $G_k = 0$. Exclusive Zero means $L_k = G_k = 0$, implying complete persistence. Exclusive Gainer means $L_k = 0$ and $G_k > 0$. Swapping Loser means $L_k > G_k > 0$. Swapping Zero means $L_k = G_k > 0$. Swapping Gainer means $G_k > L_k > 0$. Table 3 shows the various behaviors of original 64 categories. Swapping Losers and Swapping Gainers account for 87% of the study extent at both time points.
Fig. 3. Six behaviors organized as two types of swap in the rows and three types of net in the columns. Each behavior is illustrated by a pair of maps that show four pixels from two time points, where $t_1$ means the initial time and $t_2$ means the subsequent time. Black indicates the category of interest and white indicates any other category.

Fig. 4 illustrates how the algorithm’s steps are grouped into three phases. The first phase is the Exclusive Zero phase, when our algorithm aggregates pairs of Exclusive Zero categories until all Exclusive Zero categories are aggregated into one comprehensive Exclusive Zeros category, which remains as one of the three categories at the end of the algorithm. The Exclusive Zero phase determines the sequence of the categories to aggregate by selecting first the categories that have the smallest sizes.

The second phase is the Exclusive Loser and Gainer phase, when the algorithm aggregates pairs of Exclusive Gainer categories and pairs of Exclusive Loser categories. The algorithm identifies the pair of Exclusive Gainers that has the smallest $T_{ij}$ among all pairs of Exclusive Gainers, and then the algorithm identifies the pair of Exclusive Losers that has the smallest $T_{ij}$ among all pairs of Exclusive Losers. The algorithm compares the two identified pairs, and aggregates the pair that has the smaller $T_{ij}$. If two or more pairs of Exclusives have the same $T_{ij}$, then the algorithm aggregates the pair that forms the smallest sum of areas including persistence. The algorithm then looks for the next pair to aggregate. This second phase produces two categories. First is an Exclusive Losers category, which is the union of all Exclusive Losers. Second is an Exclusive Gainers category, which is the union of all Exclusive Gainers. These two categories are candidates for further aggregation in the next phase of the algorithm. Net change and swap change are constant during the algorithm’s first two phases.

The third phase of the algorithm is the swapping phase. In this last phase, our algorithm computes $T_{ij}$ for every possible pair of categories that have the same direction of net change. It then identifies the pair that has the smallest $T_{ij}$ among all losing pairs, and identifies the pair that has the smallest $T_{ij}$ among all gaining pairs. The algorithm aggregates either the identified pair of Swapping Losers or the identified pair of Swapping Gainers, whichever has the smaller $T_{ij}$. If two or more pairs of swapping categories have the same smallest $T_{ij}$, then the algorithm aggregates the pair whose sum of changes is the least. If more than one pair has the same smallest $T_{ij}$ and the same smallest sum of changes, then the algorithm aggregates the pair that has the smallest sum of areas. The newly formed category is a candidate for subsequent aggregations. Swapping Zero categories can be paired with a gainer, a loser, or another Swapping Zero category, using the same logic that this paragraph describes. The possible pairs for swapping categories can be: gainer with gainer, gainer with zero, loser with loser, loser with zero, or
Fig. 5. Net change in light shading and swap change in dark shading during 2001–2006 in terms of percent of the study extent for the behavior-based aggregation sequence.
Fig. 6. The change between 2001 and 2006 in Redland for: the 64 original categories, the 10 use-based categories and the 10 behavior-based categories. The total change is the sum of net and swap.

zero with zero. As the aggregation produces fewer categories, net change is constant while swap change shrinks by $T_{ij}$ at each step in the sequence of aggregations.

**Results**

Fig. 5 shows the entire sequence of behavior-based aggregations. Each row in Fig. 5 reports the two categories that are aggregated at each step. For example, the algorithm’s first step aggregates Sewage Treatment Plants with Penal and Correctional to form a category called Aggregated 63. The algorithm’s second step aggregates Aggregated 63 with Electric Power to form a category called Aggregated 62. The behavior-based algorithm produced 12 categories that maintain the same amount of change as the original 64 categories. The swap change shrinks slightly as the aggregation moves from 12 to 10 categories, and then the swap change shrinks substantially as the aggregation moves to fewer than 10 categories. Therefore, we focus on the behavior-based aggregation that has 10 categories, which happens to be the same number of categories as the use-based aggregation.

Fig. 6 shows the change among the original 64 categories compared to the change among 10 categories for the use-based and for the behavior-based aggregations. The use-based aggregation reduces the original net and original swap components. The behavior-based aggregation maintains the original net component and nearly maintains the original swap component.

Fig. 7 compares the use-based aggregation with the behavior-based aggregation concerning each aggregation’s 10 categories. Six singleton categories account for the majority of the change in the behavior-based aggregation. Singleton categories called Single Family Low, Plant Nurseries, and Vacant Protected Governmental Owned gain more than they lose. Singleton categories called Groves, Vacant Non-Protected Privately Owned, and Row and Field Cropland lose more then they gain. The Row and Field Cropland singleton category accounts for more change than any other singleton category and is the last category in the sequence of aggregations (Fig. 5).

Fig. 8 shows the effect of the aggregations on the maps. The original 64 categories and the behavior-based aggregation have 438 patches of change, while the use-based aggregation has 359 patches. Most of the black patches are Groves, Plant Nurseries, or Row and Field Cropland at 2001 that transitioned to Vacant Protected Governmental Owned at 2006. The red patches are mostly intra-agricultural changes, usually from Row and Field Cropland and Orchards to Nurseries.

A spreadsheet at the developer’s web site gives the original square contingency table and a detailed account of each category in terms of area and behavior at each step in the sequence of aggregations (https://sites.google.com/site/intensityanalysis). The original square contingency table shows that the largest transitions are: from Row and Field Cropland to Vacant Protected Governmental Owned, from Row and Field Cropland to Plant Nurseries, and from Groves to Vacant Protected Governmental Owned. The sum of the gross losses of the behavior-based aggregation’s six singleton categories account for 88% of all gross losses and 81% of all gross gains among the 64 original categories. Transitions among the six singleton categories are involved in 97% of the overall change in the 10-category behavior-based aggregation.

**Discussion**

**Interpretation**

We recommend that scientists examine the entire sequence of aggregations in the format of Fig. 5, to find the step(s) in the sequence that are important for a case study’s research question. The behavior-based algorithm computes the entire sequence of aggregations to three categories, but we do not necessarily recommend that scientists reduce the number of categories to three. We avoid any universal rule concerning a particular stopping point in the sequence of aggregations, because scientists should interpret results in the context of each case study. Scientists should pay particular attention to the aggregation steps where the change suddenly shrinks and to the categories that survive to later steps of the sequence. Scientists should look for categories that have relevant meaning, while realizing that the aggregation can produce potentially large but less relevant categories of Exclusive Zeroes, Gainers and Losers. Each of those three aggregated categories might consist of a collection of categories that are large in size, diverse in use, and irrelevant in change.

The behavior-based algorithm reveals a set of six detailed singleton categories and four aggregated categories for Redland. Fig. 7 shows that the six singleton categories account for more change than the four aggregated categories, thus we focus attention on the six singleton categories and their placement in the use-based aggregation. Fig. 2 shows that the use-based Agriculture category contains the three behavior-based singleton categories: Groves, Plant Nurseries and Row and Field Cropland. Fig. 7 shows that the use-based Agriculture category is involved in a small amount of change, but the three behavior-based singleton categories account for a substantial amount of change. If scientists were to use only the use-based aggregation, then they would miss the fact that three behavior-based singleton categories within the Agriculture use-based category account for substantial changes. For example, the sum of the transitions between Row and Field Cropland and Groves accounts for 2% of the study extent, which is substantial compared to the change on 17% of the study extent. Also, when the Plant Nurseries category gains, it gains the most from Row and Field Cropland and Groves. The use-based aggregation counts as persistence any transitions among these three detailed categories.

Fig. 2 shows that the use-based Undeveloped category contains the two behavior-based singleton categories: Vacant Protected Governmental Owned and Vacant Non-Protected Privately Owned. Fig. 7 shows that the use-based Undeveloped category is involved in a small amount of change, but the two behavior-based singleton categories account for a substantial amount of change. A process of acquisition by government agencies of Row and Field Cropland and Groves explains the large gain in Vacant Protected Governmental Owned. The category Vacant Non-Protected Privately Owned experiences both gross gain and gross loss, which produces a swapping
dynamic. This swapping dynamic offers evidence of an impermanence syndrome (Lopez et al., 1988), which is a process whereby owners of productive farmland cease to invest in their farmland in anticipation of development or eminent domain, thus hastening the land’s obsolescence. These insights concerning the processes of change would be more difficult to develop by examining the transitions among the 10 use-based categories or the 64 original categories.

Each particular research question will determine whether behavior-based aggregation is an important alternative to other forms of aggregation. If various behaviors exist for detailed categories that are within one broad hierarchical category, then behavior-based aggregation is important to consider. For example, scientists in FCE want to understand the process of how land becomes preserved, therefore it is helpful that the behavior-based algorithm kept separate the two singleton categories that the use-based aggregation grouped under Undeveloped (Fig. 2). Scientists in FCE are also researching how land use influences water quality, therefore it is helpful that the behavior-based algorithm kept separate three singleton categories of Agriculture, because the detailed categories of Agriculture have various influences on water quality. However, if homogeneous behaviors exist for detailed categories that are within one broad a priori category, then behavior-based aggregation is less important to consider. For example, Fig. 2 shows that the nine Open Space categories are all either Exclusive Zeros or Gainers in the behavior-based aggregation, so both the behavior-based and the use-based aggregation produce nearly the same results for those categories.

Research agenda

Our algorithm takes a stepwise approach, where each step of the algorithm aggregates a pair of categories. We are guaranteed that the algorithm maintains net change at each step. It is not clear whether our algorithm produces a universal maximum amount of change for any particular number of aggregated categories. Future research could look into this issue, but it will require thought that should avoid a brute force approach to test all possible aggregations, because the number of possible aggregations is likely to surpass computational ability. For example, if the number of categories is 64, the number of subsets is 10, and the number of categories in each subset matches the behavior-based aggregation in Fig. 2, then the number of combinations for possible aggregations is greater than $10^{12}$. Also, if the number of categories is 64, the number of subsets is 10, and the number of categories in each subset matches the use-based aggregation in Fig. 2, then the number of...
combinations of aggregations is greater than $10^{52}$. There exist many more possible aggregations that generate 10 subsets. It is not feasible to consider all possible combinations of aggregations, which is one of the main reasons why we created our algorithm. Our computer program uses Visual Basic for Applications code embedded in an Excel file, which performs the aggregation of the 64 categories of our example on a standard laptop computer in 2 min.

The next phases of this research could also examine how category aggregation can be used in conjunction with Intensity Analysis (Aldwaik and Pontius, 2012, 2013). Intensity Analysis considers a category’s change intensity, which is the ratio of the size of a category’s change to the size of the category. Our present algorithm does not consider a category’s change intensity, thus our algorithm might aggregate small categories that have intense changes early in the aggregation sequence. This could make researchers blind to such intensely changing categories. Future research could investigate whether the algorithm should consider the intensity of each category’s changes.

Furthermore, it is important to extend the ideas of this paper so they apply to cases that have more than one time interval, because the behavior-based aggregation during one time interval might be different than the behavior-based aggregation during a different time interval.

Our algorithm can be particularly helpful for datasets that have no a priori hierarchy. For example, the State of Massachusetts makes freely available data for 21-categories of land for several time points (MassGIS, 2014). The 21 categories had been aggregated from 104 categories, for which MassGIS does not give further information. Aggregation to fewer than 21 categories might be clear for some of the categories based on semantic similarity, for example there are four detailed types of residential land that could be aggregated to one broad residential category. But it is not clear how to aggregate based on semantics for some of the other categories such as Urban Open, which is defined as “Parks; cemeteries; public & institutional greenspace; also vacant undeveloped land”, or Open Land, which is defined as “Abandoned agriculture; power lines; areas of no vegetation”. We used the 21-category data to examine land change from 1971 to 1999 in the United States Long Term Ecological Research network’s Plum Island Ecosystems site, which consists of 26 towns in northeastern Massachusetts. The 21-category data show change on 15% of the study extent. Our algorithm algorithm found set of 7 categories that also show change on 15% of the study extent. Our algorithm found set of 2 categories that show change on 11% of the study extent, and nearly all of the change is net change. The algorithm highlighted Open Land as important because the algorithm kept Open Land as unaggregated until the second to last step in of the aggregation sequence. MassGIS (2014) reports one possible aggregation to two categories: developed and non-developed. The 2-category aggregation to developed versus non-developed show change on 10% of the study extent, of which 7 percentage points are net change and 3 percentage points are swap change.

We hope other scientists will apply our algorithm to professions beyond land change science. For example, economists could use the method to study labor transitions, where the study extent is a set of laborers and the categories are the labors’ occupations.
Also, demographers could use our algorithm to study migration, where the study extent is a set of persons and the categories are the persons’ countries.

Conclusions

A priori aggregation systems of detailed categories can make researchers blind to transitions among the detailed categories. This paper introduces a new algorithm to aggregate detailed categories in a strategic sequence to maintain the size of the changes among the categories between two time points. An example of land change illustrates the algorithm. Our algorithm starts by assigning each category to one of six behavior types, based on the category’s gross loss and gross gain. The algorithm can aggregate a net gaining category with a net gaining category or aggregate a net losing category with a net losing category, but never aggregates a net gaining category with a net losing category; therefore the algorithm maintains net change. Each step of the algorithm aggregates the pair of categories that maintain the maximum swap change. Results from a case study of land change show that behavior-based aggregation can reduce the number of categories from 64 to 10 while affecting the amount of temporal change trivially, while an a priori aggregation from 64 to 10 categories substantially reduces the temporal change. Our algorithm can help scientists focus on important signals of change that they might miss if they were to perform aggregation that ignores behavior. There are many potential applications, because the only input to the algorithm is a square contingency table that shows transitions among categories. Readers can obtain a computer program to perform the aggregation by writing to the first author or by visiting https://sites.google.com/site/intensityanalysis/.

Acknowledgements

The United States’ National Science Foundation (NSF) supported this work via its Coupled Natural Human Systems program via grant BCS-0709685 and its Long Term Ecological Research program via grants OCE-0423565, OEC-1058747 and OEC-1238212 for Plum Island Ecosystems and OCE-0620959 for Georgia Coastal Ecosystems. NSF supplied additional funding through a supplement grant entitled “Maps and Locals” via grant DEB-0620579. Professors Colin Polsky, Samuel Ratick and John Rogan were members of the dissertation committee of the first author, whose dissertation presented the first version of this manuscript. Any opinions, findings, conclusions, or recommendations expressed in this article are those of the authors and do not necessarily reflect those of the funders. Clark Labs facilitated this work by creating the GIS software Idrisi®. Clark University students and anonymous reviewers supplied constructive feedback that helped to improve this paper.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jag.2014.09.007.

References


