Characterizing Regions through Content Classification

Martin Tomko[1], Ross S. Purves[1]

[1]Department of Geography, University of Zurich
martin.tomko@geo.uzh.ch, ross.purves@geo.uzh.ch

INTRODUCTION

For information retrieval tasks, georeferenced information objects can be annotated with references to spatial objects contained within the corresponding geographical footprint. Not all the spatial objects, however, describe the essential attributes characterizing the region.

For example, Venice and Amsterdam are cities often characterized by references to the network of canals that criss-crosses them. While there are shops or post offices in these cities, the canals set them apart from their European peers. Given that it is possible for humans to effectively and evocatively characterize regions we can ask what the properties of such descriptions are, and can they be formalized in order to automatically describe geographic regions?

We present a computational method to determine the descriptive prominence of categories of spatial objects in a given region and select the most prominent ones for inclusion in the characteristic description of the region. Our hypothesis is that we can identify descriptively prominent categories of spatial objects by studying the variation of their distribution in space. A definition of descriptive prominence is introduced, adapted to the selection of spatial objects of both rare and frequently occurring categories which may contribute to the character of the region.

The method is demonstrated on three datasets of points of interest (POI), with a fourth artificial dataset used as benchmark (Tomko & Purves, 2008). The method reduces the number of categories describing regions significantly ($p<0.001$). We illustrate the results qualitatively with characterizations sourced from narrative text.

CHARACTERISTIC DESCRIPTIONS OF REGIONS

Descriptions are a form of verbal communication intended to identify an object (referring expressions) or characterize its properties (attributive expressions)(Ludlow, 2008). A hypothetical full attributive description of an

object contains references to all available properties of the object. In the case of a region, this would imply an overloaded description with references to all spatial objects found within. In typical human-generated descriptions of regions, only references to prominent spatial objects or their categories are included. We call such descriptions *characteristic geographic descriptions*.

In order to automatically generate characteristic descriptions of regions, we select references to prominent objects or categories of spatial objects. While the prominence of individual spatial objects can be determined by, for example, assessing their visual properties (Nothegger et al., 2004), no method is currently available to identify prominent categories of spatial objects.

Descriptively prominent categories of spatial objects may be defined by application of the concept of contrast from background as categories of spatial objects that are over-represented in the region described. We define over-representation of a category of spatial objects as follows:

> Category c is relatively over-represented in region A^x if and only if the likelihood of occurrence of an object of category c in the region A^x is higher than the likelihood of its occurrence in the containing region A^{x-1}.

Under-representation is then the negation of over-representation.

METHOD: IDENTIFICATION OF REFERENCES FOR CHARACTERISTIC GEOGRAPHIC DESCRIPTIONS

Let A^1, A^2, A^3 be hierarchically nested geographical regions, where A^1 is the largest and A^3 is the smallest region. We assess the descriptive prominence of a category of spatial objects by comparing the frequency of its occurrence in these three hierarchically nested regions. The combinations of over or under-representation of a category at each granularity can be typified into eight cases (Tab. 1), where categories of spatial objects descriptively prominent for the characterization of region A^1 belong to Cases 1, 3, 5 and 7. The construction of characteristic descriptions requires the classification of categories of objects found in the regions described under the eight cases and the filtering of under-represented categories out of the final annotation set.

Tab. 1: Occurrence cases. + stands for over-representation, - for under-representation.

Region	A^3 *(small)*		A^2 *(medium)*		A^1 *(large)*	
Case	+	-	+	-	+	-
1	X		X		X	
2		X		X		X
3		X	X		X	
4	X			X		X
5	X			X	X	
6		X	X			X
7		X		X	X	
8	X		X			X

METHOD TESTING AND RESULTS

The method outlined above was tested on POI datasets covering parts of the UK (Cardiff), North Holland and Switzerland (Zurich). For statistical assessment, regions A^3 were randomly generated in the complete space covered by the individual datasets (regions A^1), while regions A^2 were represented by the partitioning of the regions A^1 into proximal regions around settlement centroids.

We observe that by filtering out under-represented categories of spatial objects from the reference sets, we achieve a statistically significant reduction of the references for inclusion in characteristic descriptions of regions. The more unique a category is, the more likely it is to be preserved, while the abundant categories are frequently retained, allowing for references to common and likely typical categories of spatial objects.

We illustrate the results with an example of characteristic description of Hadfield Road, Cardiff, including a human-generated literary narrative:

- All POI categories in region, with classification: *{PetrolStation [Case 3], Car Dealer [Case 1], Rent a Car Facility [Case 8]}.*
- Descriptively most prominent category: *{Car Dealer }.*
- Human-generated description: *"Auto land. Hadfield is car showrooms run together on the me-too principle from Plexiglas beginning to crystal laminated end." (Finch, 2004).*

CONCLUSIONS

The proposed method for constructing characteristic geographic descriptions was demonstrated only on POI datasets and shows plausible statistical results, further assessed in a qualitative manner. An evaluation based on subject testing using georeferenced photographs will follow. Further work will integrate references to prominent categories of linear or areal spatial objects, as well as references to visually salient spatial objects.

ACKNOWLEDGMENTS

The research reported is part of the project TRIPOD supported by the EC under contract No. 045335.

REFERENCES

Finch, P. (2004). Hadfield Road. In *Real Cardiff* Bridgend, Wales, UK: Seren Books.

Ludlow, P. (2008). Descriptions. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring Edition ed.): Standford University.

Nothegger, C., Winter, S., & Raubal, M. (2004). Computation of the Salience of Features. *Spatial Cognition and Computation, 4*(2), 113-136.

Tomko, M., & Purves, R. S. (2008). *Categorical Prominence and the Characteristic Description of Regions* Paper presented at the Semantic Web meets Geospatial Applications, held in conjunction with AGILE 2008, Girona, Spain.