Analysis of community-contributed space- and time-referenced data (example of flickr photos)

Gennady Andrienko, Natalia Andrienko, Peter Bak, Slava Kisilevich, Daniel Keim

University of Bonn and Fraunhofer Institute IAIS, University of Konstanz

ABSTRACT

Space- and time-referenced data published on the Web by general people can be viewed in a dual way: as independent spatio-temporal events and as trajectories of people in the geographical space. These two views suppose different approaches to the analysis, which can yield different kinds of valuable knowledge about places and about people. We present several analysis methods corresponding to these two views. The methods are suited to the large amounts of the data.

1 INTRODUCTION

In the age of Web 2.0 more and more people publish various kinds of contents on the Web. Some kinds of contents have spatial and temporal references, for example, the photos in flickr (www.flickr.com) linked to the places where they were taken and supplied with the dates and times of the shots. Such data can serve as a source of knowledge about the places and about the interests, behaviors, and mobility of the people. The knowledge may be valuable for local administrations, tourist services, advertising agencies, and other organizations. However, the data are not easy to analyze. One of the problems is the huge number of entries, which calls for scalable computational techniques. At the same time, the involvement of a human analyst, who perceives spatial and temporal patterns and gives them meaning, is essential.

The community-contributed Web entries having spatial and temporal references can be viewed, on the one hand, as independent spatio-temporal events. On the other hand, the entries made by the same person can be considered as a trajectory of this person in the geographical space, which tells something about the movement and behavior of this person. The whole dataset can be viewed as a set of trajectories of multiple people. These two views suppose different approaches to the analysis, which can yield different kinds of knowledge. Another example of data that can be viewed in such a dual way is data about mobile phone calls. We call this class of data *event-based movement data*.

Analysis of event-based movement data is a relatively new research topic. A series of papers have been published by Girardin and co-authors (e.g. [2] and [3]). They analyze concentrations and movements of tourists at the scales of a city (e.g. Rome) and a geographical region (e.g. central Italy) using Flickr photos and, in some studies, mobile phone calls made by the tourists in the same time periods. Concentrations are shown on heat maps produced by dividing the area of interest by a rectangular grid and counting the number of photos in every grid cell. Interpolation techniques are used in order to smooth the visualization. The movements are visualized in an aggregated form by means of flow maps, where predefined places are connected by lines with the widths proportional to the numbers of tourists that moved between the places. In [1], another research team uses geo-annotated Flickr photos to find concentrations of activity and most popular places on Earth. For that they use a non-parametric MeanShift clustering algorithm. The authors have also used the temporal information available from the photos to generate a few example maps showing the movements of photographers. The trajectories are represented by lines drawn on the maps in a semi-transparent mode; however, no further analysis is made.

2 OUR APPROACH AND METHODS

We take a systematic approach to the analysis of event-based movement data. We define possible types of analysis tasks related to the views of the data as events and as trajectories. We distinguish space-centered tasks, where the data are used to study the properties of the space and places, from agent-centered tasks targeting at the properties and behaviors of the people (in general, moving agents). For the tasks defined, we try to find or develop appropriate methods.

	Space-centered	Agent-centered
Events	To be completed…	
Trajectories		

For space reasons, only two methods are briefly introduced here. The example data subset consists of photos made in the area of Atlantic City. The methods do not take into account the contents of the photos but only their spatial and temporal references.

2.1 Spatio-temporal analysis of clusters of events

The Growth Ring Maps technique supports the exploration of the frequencies and temporal patterns of events occurring in the same places. For the flickr dataset, we defined the significant places on the basis of the density-based spatial clustering of the events. For each place, we represent the photos made in it by pixels placed around the central point in an orbital layout according to the times of taking the photos: the earlier the photo was made, the closer the pixel is to the central point. Color-hue is used to map semantic properties of the events or places. Thus, seasonal differences in visiting the places may be investigated by mapping the seasons to four distinct colors (winter-white, spring-green, summer-red, and autumn-blue). The resulting map shows simultaneously the intensity of taking photos at different locations and the seasonal differences. Figure 1 exposes salient distinctions among the seasonal patterns of taking photos in different places of Atlantic City. In the most popular places, the photos were taken all the year around but more photos were made in summer and autumn. In the other places, we see a clear prevalence of one season.

Figure 1. Growth Ring Maps showing the seasonal differences of taking photos for Atlantic City.

2.2 Analysis of flows

For this analysis, we build flow maps showing aggregated moves between places, i.e. how many people have moved from one place to another. There are two major differences from what is described in [2] and [3]. First, our flow maps show not only the amounts but also the directions of the movement by special half-arrow symbols. It is easy to see whether the movements between two locations are one- or two-directional and, in the latter case, whether one of the directions prevails over the other. Second, we do not use any predefined places but divide the territory into appropriate compartments on the basis of clustering of the positions from the trajectories. By varying the clustering parameter (specifically, the desired cluster radius), we can do the analysis at different spatial scales. An example is presented in Figure 2. We have divided the sequence of photos made by each person into subsequences, or sessions, assuming that a time interval of 8 hours or more between two photos means the beginning of a new session. The sessions have been treated as trajectories. First, we tessellated the territory into bigger areas using clustering with the desired radius 2.5km. The upper map fragment shows the flows between the areas. By means of interactive filtering, we have hidden minor flows. We can see that people typically do not make long walks during one photo session. Almost all movements occur in a limited area on the northeast of the shown territory.

The lower map fragment represents the movements in the central part of the city aggregated at a smaller spatial scale (the cluster radius was 200m). By filtering, we have removed the minor flows. We can see movements along the seafront and movements from the street to two piers and back.
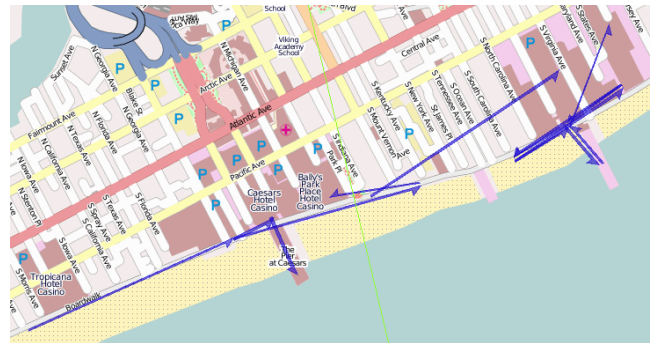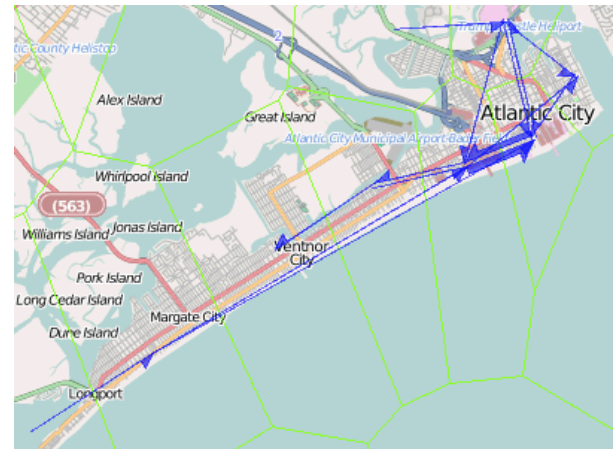

Figure 2. The flows of the photographers in Atlantic City at two different spatial scales.

3 DISCUSSION

In most of the existing approaches to analysis of movement data it is assumed (in most cases implicitly) that the available position records represent continuous trajectories and, hence, intermediate positions can be obtained by means of interpolation between known positions. However, there are many cases when the position records are temporally sparse and irregular, which means that the data cannot be handled in this way. This kind of data can be called event-based movement data. We suggest a set of visual analytics methods combining computational techniques with interactive visual displays to support the analysis of such data. The methods are oriented to different classes of analysis tasks and enable the analyst to discover different kinds of patterns in the data, as is demonstrated by example of flickr photos.

REFERENCES

[1] Crandall, D., Backstrom, L., Huttenlocher, D., and Kleinberg, J. 2009 Mapping the World's Photos. International World Wide Web Conference.

[2] Girardin, F., Calabrese, F., Dal Fioro, F., Ratti, C., Blat, J. 2008. Digital Footprinting: Uncovering Tourists with User-Generated Content. Pervasive Computing, IEEE, 7, 36-43

[3] Girardin, F., Dal Fioro, F., Ratti, C., Blat, J. 2007 Understanding of Tourist Dynamics from Explicitly Disclosed Location Information. In 4th International Symposium on LBS and Telecartography, Hong-Kong, China.