Chapter 7

Mining Geospatial Knowledge on the Social Web

Suradej Intagorn
USC Information Sciences Institute, USA

Kristina Lerman
USC Information Sciences Institute, USA

ABSTRACT

Up-to-date geospatial information can help crisis management community to coordinate its response. In addition to data that is created and curated by experts, there is an abundance of user-generated, user-curated data on Social Web sites such as Flickr, Twitter, and Google Earth. User-generated data and metadata can be used to harvest knowledge, including geospatial knowledge that will help solve real-world problems including information discovery, geospatial information integration and data management. This paper proposes a method for acquiring geospatial knowledge in the form of places and relations between them from the user-generated data and metadata on the Social Web. The key to acquiring geospatial knowledge from social metadata is the ability to accurately represent places. The authors describe a simple, efficient algorithm for finding a non-convex boundary of a region from a sample of points from that region. Used within a procedure that learns part-of relations between places from real-world data extracted from the social photo-sharing site Flickr, the proposed algorithm leads to more precise relations than the earlier method and helps uncover knowledge not contained in expert-curated geospatial knowledge bases.

INTRODUCTION

When disaster strikes, people are increasingly turning to social media sites to reach out to friends and family, post information, including images and videos, about current conditions, and receive updates about shelters and safe harbors. Humanitarian relief community could potentially integrate this data with geo-spatial information contained in maps, satellite imagery and news reports, to monitor the unfolding situation, assess damage and help coordinate relief efforts. As an example, images posted on the social photo-sharing site Flickr both before and after a tornado devastated a
town, could be combined with eyewitness accounts and missing persons reports on the microblogging service Twitter to create a detailed view of the affected area and its population. First-hand accounts on Twitter could then be used to monitor the availability of shelter and critical supplies, such as fresh water. The challenge, however, is to link places people talk about in their posts to the actual geo-spatial entities, since ordinary people are highly unlikely to use terms from a predefined geo-spatial vocabulary, or may refer to places that are not formally defined, such as, neighborhoods and landmarks.

We address this problem by automatically mining social media content to learn about places and relations between them. In addition to creating rich content in the form of text documents, images, and videos on the Social Web sites such as Flickr and YouTube, people often annotate content with keywords, called tags that they use to label and categorize content, as well as geographic coordinates, or geo-tags. Although social metadata lacks a controlled vocabulary and predefined structure, it reflects how a community organizes knowledge, including geospatial knowledge. A corpus of social metadata created by large numbers of people can be mined to reveal concepts (Plangprasopchok, 2004) including places and relations between them (Keating, 2005). Community-generated knowledge that is automatically extracted from social metadata can complement expert-curated geospatial knowledge (Keating & Montoya, 2005; Kavouras et al., 2006), such as Geonames (http://geonames.org) or Yahoo! GeoPlanet (http://developer.yahoo.com/geo/geoplanet/).

Community-generated knowledge is more likely to stay complete and current, since it is learned from metadata that is distributed and dynamic in nature (Golder & Huberman, 2006). It is also more likely to reflect colloquial folk knowledge that people use to talk about places.

Recently we proposed a method for aggregating geo-tagged data created by thousands of users of the social photo-sharing site Flickr to learn places and relations between them (Intagorn et al., 2010). The method represents a place by the coordinates of the geo-tagged images Flickr users labeled with the place name and uses geospatial subsumption to learn relations between places. Our key challenge is to efficiently and accurately represent places. In the original work, we used convex hulls to represent places, but found they did a poor job, since places were often concave. To address this problem, we present a simple, computationally efficient algorithm to find a possibly concave contour of a planar shape. Our method starts with a bounding box that subsumes all points and gradually erodes it until the boundary converges to a polygon that best represents that shape. We evaluate the method on data set consisting of US zipcodes. We then apply it to learn the boundaries of places extracted from social metadata. We show that the new method enables us to learn more precise relations between places using geospatial subsumption. Some of what we learn includes novel relations not found in the formal directories, for example, that Wild Animal Park is in San Diego. While not technically correct, such expressions of folk knowledge are still quite useful.

SOCIAL METADATA AND THE CHALLENGES OF LEARNING FROM IT

Tagging has become a popular method for annotating content on the Social Web. When a user tags an object, for example, an image on Flickr, she is free to select any keyword from an uncontrolled vocabulary to describe it. In addition to tags, some social Web sites, such as Delicious, and Flickr, provide the ability for users to organize content hierarchically. While the sites do not impose any constraints on the vocabulary or the semantics of the hierarchical relations, in practice users employ them to represent subclass relationships (‘paris’ is a kind of ‘city’) and part-of relationships
Related Content

Overpopulation and Its Association With Natural Disasters: A Case Study of Indian Tsunami, 2004
www.igi-global.com/chapter/overpopulation-and-its-association-with-natural-disasters/207632?camid=4v1a

Real-Time Data Visualisation in Collaborative Virtual Environment for Emergency Management
www.igi-global.com/article/real-time-data-visualisation-in-collaborative-virtual-environment-for-emergency-management/108892?camid=4v1a

Equipment Distribution for Structural Stabilization and Civilian Rescue
www.igi-global.com/article/equipment-distribution-structural-stabilization-civilian/53233?camid=4v1a

Lessons from Major Incidents Influencing and Influenced by Telecoms Failures
www.igi-global.com/chapter/lessons-from-major-incidents-influencing-and-influenced-by-telecoms-failures/90722?camid=4v1a