Social Interactions over Geographic-Aware Multimedia Systems

Roger Zimmermann
School of Computing, Interactive & Digital Media
Institute, National University of Singapore
Singapore 117417
rogerz@comp.nus.edu.sg

Yi Yu
School of Computing
National University of Singapore
Singapore 117417
yuy@comp.nus.edu.sg

ABSTRACT

User-centric Internet multimedia scenes challenge us to discover more interesting events and topics based on matching users' needs associated with personal preferences, geographic interests and social norms. Geotagged multimedia contents from online social sites, *e.g.*, Flickr and Twitter, provide large volumes of data about many given locations. Hence, location is one of the most important user-generated contexts and contains rich information about an individual's interests and behavior. Location-based social multimedia streams (*e.g.*, tweets, videos, images) can provide us with socially complementary information to predict users' needs.

By making use of geo-tagging, a cohesive set of social multimedia streams can be published to facilitate a more accurate analysis of user-centric big data information and further assess user tastes based on location activities. This tutorial delivers not only a better understanding of the basics of location-aware contextual descriptions and its relations to social multimedia scenes, but may also serve to highlight relationships that can be collaboratively applied to multimodal retrieval and recommendation technology.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval; I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis—*Sensor Fusion*

Keywords

Geographic sensor information; location-based services; social interactions; mobile videos; location recommendations

1. BACKGROUND

Today the web is able to lively visualize modern life since we can upload our photos, audios, videos, views and blogs onto the Internet and share them with other people on various social media sites such as YouTube, Flickr, Foursquare, Twitter and Weibo. Images and videos are especially important because they are generated, tagged, and tweeted by

users. With an overwhelming amount of social media content on the Internet, it is difficult to find what users are really interested in through navigating massive multimodal collections. For example, a search for "Merlion" may result in the return of thousands of social media items caused by diverse users' query intentions at different social websites.

In general many similar media content items appear on the Internet. To improve the quality of retrieval and recommendation systems, some extra information, such as geo-sensor information captured by mobile devices, can be considered to associate specific social media streams with locations and further predict what kind of topics over social interactions in multimedia streams are most important for users. This poses a number of interesting questions: (1) At a specific location, different media streams may have different views. How are they respectively influencing video popularity over social media streams? (2) Can these media streams be associated with each other to retrieve and recommend user-generated videos (UGV) based geo-sensor information? (3) Does geographic locality of users' behaviors and biases have an impact on the topics and events of video retrieval and recommendation? (4) What principles can be applied to incorporate social media streams into multimedia collection annotations according to geo-tagging information? These issues will be discussed in this tutorial.

2. MOTIVATION

Location-based services enable some social media streams to obtain users' places or coordinate information. Geo-location can represent several unique and semantic user contexts (*e.g.*, physical world, culture, behavior) that complement the performance of discovering and retrieving other social media by annotating geographic information. Basically, geographic metadata is utilized as an index to find relevant contents within area [4, 3]. Moreover, crowdsourcing (*e.g.*, leveraging the Foursquare API) as an affordable and effective tool can be used to annotate geographic metadata with meaningful categories [1, 5] to make them more informative for geographic-aware personalized location recommendations [1] and entertainment generation [5]. Although many researchers have investigated the association of additional visual features [2] to find relevant content, the performance of geographic multimedia retrieval and recommendation is still limited, especially for commercial big data sets on emerging social platforms.

We conjecture that geographic-aware social media streams provide a better unified set to predict the live needs of users. Interesting events and topics across various social multime-

dia streams are varying at a specific location, which results in trends constantly changing (*e.g.*, images, videos, messages and tweets on social location services). Fortunately, geographic-based, aggregated social media streams are able to serve as a diverse, fresh and robust repository to predict the desired content for a particular user. It is appealing to model users' needs associated with personal preferences, geographic interests and social promotions by leveraging geographic-aware multimodal social media streams. In general we would like to consider various geographic-aware aggregated multimedia streams to address how and why three intuitive aspects (personal preference, geographic interest and social promotion) respectively or integrally influence a particular user's need findings. We investigate topics and methods for handling interesting user-generated video (UGV) discovery and multimodal retrieval and recommendations over geographic-aware social media streams.

3. OBJECTIVES

Inspired by the diversity of location-based, user-centric social multimedia data sources, this tutorial describes emerging users' needs and multimedia applications based on locations, reviews the state-of-the-art in social geo-tagged multimedia systems, surveys several intuitions of geo-aware social media streams, and shares some preliminary results obtained through crowdsourced annotations of a richer social multimedia dataset. We also will introduce different evaluation benchmarks and address some promising trends related to geographic-aware social multimedia recommendations. A more detailed outline is provided in the next section.

This tutorial is customized for intermediate-level audiences consisting of researchers who are quite familiar with indexing of geo-tagged images and videos or multimedia content analysis but are not familiar with social context analysis, and either intend to model the interactions among mobile people and multimedia events or wish to get involved into location-based topic discovery in multimedia systems. We hope to broaden the understanding of location-aware user-centric big data in multimedia systems beyond simple location-based contextual information. We plan to distribute materials including course slides, relevant links, online resources and source codes.

4. TUTORIAL OUTLINE

The tutorial is organized into four main modules. First, we introduce geo-tagging techniques as well as related features. We then discuss several issues of discovering the most important events and topics over multimedia streams. Next, we provide an extensive overview of location-based research trends in various multimedia scenarios covering a broad range of topics and methods. We also show demonstrations of location-based multimedia retrieval and recommendation systems.

Module One aims to give an understanding of the advantages and disadvantages of today's video tag annotations based on a rich set of meta-data obtained from GPS and compass sensors. Furthermore, participants will also be provided with an overview of geo-tagging techniques and they will learn how geo-tagging does differentiate itself from content-based or crowdsourcing-based video tagging. As a demonstration example, we introduce the recent development of the *GeoVid* georeferenced video smartphone app

and its application (properties of viewable scene, geospatial video recording for Android and iOS, assigning tags to visible objects, association of tags with video segments).

Module Two aims to give an understanding of why we need to incorporate geographic-aware social media information (*e.g.*, mitigating cold-start for new individuals and data sparseness) to predict multimedia scenes desired by users and why we need to consider three aspects (personal preference, geographic interest and social promotion) to find the most interesting events and topics of videos. We also introduce recent geo-tagged multimedia applications and techniques (*e.g.*, land-use classification, geo-tagged image retrieval, preference-aware location recommendation). We further provide geo-tagged metadata to address how we are able to incorporate social media information to find the most interesting videos fitting a particular user's needs.

Module Three focuses on location-enabled social topics and interactions around multimedia contents and streams. We introduce a set of basic components (*e.g.*, personalized location, geographic-aware local interest, geo-category), topics (*e.g.*, social knowledge learning, personal preference mining, preference-based scoring, location rating, geographic distribution of social views), methods (*e.g.*, SVMhmm, iterative learning model, sparse additive generative model) and evaluation matrices (*e.g.*, recall, precision, mean reciprocal rank). We also provide some novel examples to address challenging issues, *e.g.*, how to link social media streams to multimedia contents over geographic information, and will share our studies with the participants.

Module Four guides the participants through various resources covering software tools, public datasets and crowdsourcing methods used in the research communities of multimedia, collaborative recommendations, topic tracking and social data mining & discovery. Various multimedia datasets have been collected and described as part of the ACM Multimedia Systems Dataset Track to help explore various application needs. We also demonstrate some typical location-based multimedia tasks according to different users' needs and different multimedia scenarios.

Acknowledgments

5. REFERENCES

[1] J. Bao, Y. Zheng, and M. F. Mokbel. Sensor Information Location-based and Preference-Aware Recommendation Using Sparse Geo-Social Networking Data. In *ACM SIGSPATIAL GIS*, pages 199–208, 2012.

[2] J. Kamahara, T. Nagamatsu, and N. Tanaka. Conjunctive Ranking Function using Geographic Distance and Image Distance for Geotagged Image Retrieval. In *ACM Geo Workshop*, pages 9–11, 2012.

[3] H. Ma, R. Zimmermann, and S. H. Kim. HUGVid: Handling, Indexing and Querying of Uncertain Geo-tagged Videos. In *ACM SIGSPATIAL GIS*, pages 319–328, 2012.

[4] Z. Shen, S. Arslan Ay, S. H. Kim, and R. Zimmermann. Automatic Tag Generation and Ranking for Sensor-rich Outdoor Videos. In *ACM Multimedia*, pages 93–102, 2011.

[5] Y. Yu, Z. Shen, and R. Zimmermann. Automatic Music Soundtrack Generation for Outdoor Videos from Contextual Sensor Information. In *ACM Multimedia*, pages 1377–1378, 2012.