Tourism Knowledge Discovery in Social Multimedia

Symeon Papadopoulos
Information Technologies Institute, Greece

Adrian Popescu
Vision and content Engineering Laboratory, France

Ioannis Kompatsiaris
Information Technologies Institute, Centre for Research and Technology Hellas, Greece

INTRODUCTION

Tourism has been one of the first domains of human activities, where the role of multimedia has been crucial as a means to capture and archive personal experiences. Recently, the wide adoption of digital cameras and the popularity of social sharing platforms led to the proliferation of travel-centric user-generated content (UGC) on the Web. Currently, there is a wealth of travel-centric user-contributed information and content on the Web that adds significant value to "traditional" sources of tourism knowledge and content (e.g. official travel guides):

- Images and videos captured by casual tourists during their trips and shared through social networks (Flickr, Panoramio, YouTube).
- Blog articles describing the experiences of travelers.
- Structured tourism information, e.g. Wikipedia infoboxes on popular landmarks or areas.
- Tourism behavioral data, e.g. Foursquare check-ins.
- Explicit feedback and travelers' impressions, e.g. comments in popular travel sites such as TripAdvisor and booking.com.

As a result of the wealth and diversity of online tourism-related information and content, it has been recently possible to employ data mining, multimedia analysis and knowledge discovery approaches on the tourism domain for refining the information about travel destinations, e.g. by enriching the representation of existing *Points of Interest* (PoIs) and discovering new ones, and for gaining new insights into the visiting patterns and traits of travelers. Such insights have also made possible the development of innovative services that could greatly enhance the travel experience, for instance, PoI recommendation (Popescu & Grefenstette, 2011) or *itinerary* creation (Tai et al., 2008). Furthermore, tourism knowledge discovery has proved to be invaluable for tourism professionals by providing insights regarding the visiting patterns and preferences of visitors without the need to resort to costly questionnaire-based surveys. This article summarizes several important research problems and novel approaches in the emerging domain of *tourism knowledge discovery* from social media content and online user activities.

BACKGROUND

Here, we first motivate the presented work by highlighting the importance of related technologies for the tourism industry as seen in recent research. Then, we describe the main information and content sources involved in the process of multimedia knowledge discovery. Finally, we provide definitions for each of the research problems that are further discussed in the article.

Significance of ICT, Web, and Social Media for the Tourism Industry

The Internet is one of the most influential technologies affecting travelers' behaviour. Prospective travelers have nowadays access to vast amounts of information by tourism organisations, private enterprises and increasingly by other tourists thanks to the emergence of social media. As pointed out by a thorough survey on the use of ICT in tourism (Buhalis & Law, 2008), one of the main pertinent challenges is the prevalence of information overload and the availability of too many choices for prospective travelers. This situation highlights the need for effective recommendation technologies that greatly benefit from the availability of multimedia knowledge discovery methods. The same study emphasizes the role of multimedia as a means of tangible imagery and experience to travel planners.

According to another survey on tourism innovation (Hjalager, 2010), the emergence of Web, search engines and social media are considered a key driver of management innovation, and the use of smartphones for delivering sophisticated location-based applications is a prominent example of a technology push factor that could transform the respective business models. A prototype mobile multimedia application, built on top of the results of social multimedia data mining, is presented later in this article.

Xiang and Gretzel (2010) conducted a user study to confirm the importance of social media in travel information search and planning. According to it, links to social media sites represent at least 10% of travel search results, of which a large portion refers to virtual communities, consumer review sites, blogs and media sharing sites. The increasing importance of blogs and Web 2.0 platforms for tourism information exchange was also highlighted in an independent review (Schmallegger & Carson, 2008) and associated with business functions such as communication, promotion, product distribution, management and research.

All aforementioned studies confirm the significance of the Internet, social media platforms, online communities, and user-generated content for the support of innovation in the tourism industry. In this article, we focus on the research problems that are necessary to tackle to leverage the value hidden in UGC published in social media, and propose a research framework that links ongoing research work with these problems.

Tourism Information and Content Sources

Knowledge discovery is typically based on three types of tourism data: (a) multimedia collections shared by tourists online, (b) behavioral data collected from location-based applications, and (c) gazetteers of geographical entities.

Multimedia collections offer a starting point for tourism knowledge discovery processes. Such collections consist of content items (images, videos) and accompanying metadata. Each item is typically contributed by an online account (owner/contributor) and is associated with text metadata (title, description, tags), the location where it was captured (e.g. latitude-longitude coordinates) and the timestamp of its creation. Several multimedia mining methods also make use of the content visual information through a set of features extracted by processing the visual information (pixels) of content items.

Besides multimedia data, social media platforms, such as Foursquare, result in *behavioral data* in the form of *check-ins*, i.e. real-time reports of users' location, typically associated with PoIs, such as hotels, restaurants, etc. Behavioral data often includes opinions or scores.

A *gazetteer* (geographical database) is a set of PoIs or geographic features. Each gazetteer entry is associated with a name, a location, a geographic category and a set of category-dependent properties (such as population for a city or country, popularity rank scores for a tourist PoI, etc.).

Research Problems

Tourism knowledge discovery is related to several challenging research problems that recently became more approachable thanks to the wide availability of online multimedia content and accompanying metadata and the commoditization of high-performance computing infrastructure. More specifically, these problems include: multimedia geotagging, PoI discovery, itinerary mining, PoI recommendation, and place summarization.

- **Multimedia Geotagging:** Given a media item, with no available location information, provide an estimate of its location. In its most challeng-

8 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/tourism-knowledge-discovery-in-social-multimedia/112800

Related Content

Self-Awareness and Motivation Contrasting ESL and NEET Using the SAVE System

Laura Vettraino, Valentina Castello, Marco Guspiniand Eleonora Guglielman (2018). *Encyclopedia of Information Science and Technology, Fourth Edition (pp. 1559-1568).*

www.irma-international.org/chapter/self-awareness-and-motivation-contrasting-esl-and-neet-using-the-save-system/183870

Why It Is Difficult to Disengage From Facebook

Sonda Bouattour Fakhfakh (2018). *Encyclopedia of Information Science and Technology, Fourth Edition (pp. 7190-7199).*

www.irma-international.org/chapter/why-it-is-difficult-to-disengage-from-facebook/184415

Forecasting Techniques for Energy Optimization in Buildings

Fernando Mateo, Juan José Carrasco, Abderrahim Sellami, Mónica Millán-Giraldo, Manuel Domínguez, Ignacio Díazand Emilio Soria-Olivas (2015). *Encyclopedia of Information Science and Technology, Third Edition (pp. 967-977).*

www.irma-international.org/chapter/forecasting-techniques-for-energy-optimization-in-buildings/112490

Studying Virtual Work in Teams, Organizations and Communities

Daniel Robeyand Leigh Jin (2004). *The Handbook of Information Systems Research (pp. 150-165).*

www.irma-international.org/chapter/studying-virtual-work-teams-organizations/30348

Hybrid TRS-PSO Clustering Approach for Web2.0 Social Tagging System

Hannah Inbarani H, Selva Kumar S, Ahmad Taher Azarand Aboul Ella Hassanien (2015). *International Journal of Rough Sets and Data Analysis (pp. 22-37).*

www.irma-international.org/article/hybrid-trs-pso-clustering-approach-for-web20-social-tagging-system/122777