Assessing the usefulness of online image annotation services for destination image measurement

Lyndon Nixon
MODUL University Vienna, Austria
lyndon.nixon@modul.ac.at

Abstract

This research note reports on an initial exploration of the usefulness of online image annotation services for the measurement of destination image. Destination Marketing Organisations (DMOs) today, while actively data mining textual content for insights into visitor sentiment towards their destination or the most popular topics or themes of visitors at that destination, increasingly face usage of digital imagery or videos - yet non-textual content is not as easily 'understood' by machines to provide the same insights. The recent emergence of online services for image annotation might be of value to DMOs but to the best of the author's knowledge no evaluation of their usefulness has been made in the touristic domain. We present here initial results which indicate the progress that image annotation services still need to make before DMOs could use them in destination image measurement.

Keywords: media mining, media analysis, media annotation, visual analysis, concept detection, image annotation, destination image, tourism intelligence.

1 Introduction

Digital travellers are more likely today to learn about or check out potential destinations through photos on Instagram, videos on YouTube or visual pins on Pinterest than to read text-based travel blogger entries, travel guides or DMO websites. The shift in digital consumer behavior towards (audio)visual content on the Web raises a new challenge for tourism stakeholders who traditionally have performed data analysis for market research and prediction based on textual and statistical data. Insights into how destinations and tourism offers are being presented to online consumers will only capture the whole story if the (audio)visual content can be analysed and understood in the same way as today's tourism intelligence solutions can perform with text. Modern advances in computational understanding have enabled significant progress in computer systems that can accurately identify concepts in visual content and label frames according to emotional characteristics, objects and events. Such powerful visual annotation capabilities are even made available publicly via Web services, meaning that functionality that has long been only accessible to very few based on highly complex and expensive computer systems is now a possibility for any business who identifies a business need for it. DMOs and other tourism organisations could benefit from the use of state of the art media annotation in order to introduce or extend their text-based tourism intelligence capabilities.

2 Related work

Tourism marketers have long been interested in the "destination image" that their audience has of the destination and how to influence that image through their own marketing content. The analysis of non-textual content for tourism marketing began

with tourist photographs taken at a destination. With the Web and social media providing free public and global distribution channels for content, combined with the ease of creation of digital image and video assets, tourism media about destinations is now being created at a huge scale, by a very large number of smaller channels of travel blogs and individual travellers. Since travellers now also increasingly use social networks as a source of information about destinations (Xiang & Gretzel, 2010), more recent studies have turned to destination image from online media. Stepchenkova and Zhan (2013) compared DMO and Flickr photos along 20 destination attributes, constructing maps of the projected and perceived images of Peru. An aggregated destination image can be formed following the procedure by calculating the frequency of occurrence of destination image attributes in the sample (Stepchenkova and Li, 2012; 2014). Fatanti and Suyadnya (2015) looked at how Instagram creates a tourism destination brand, analysing the promotional value of Instagram through "photo elicitation interview (PEI)". Tourism research has looked at the use of Instagram in destination marketing (e.g. Hanan & Putit, 2014). Nixon (2017) has tested if Instagram content can positively influence a person's perception of a destination.

3 Motivation

While tourism research has only partially covered the insights that images or videos could provide and generally examined small numbers of media assets which have been manually annotated in advance by human experts, advances in computer vision (what concepts computers can “see” in an image) offer tourism organisations the possibility to accurately annotate their own images as well as images being shared by visitors about their destinations, in order to gain deeper and valuable insights into the common topics and interests of visitors with respect to their destinations, as well as adapt their own (media) marketing campaigns appropriately.

For the initial evaluation, we chose Vienna Tourism whose account at www.instagram.com/viennatouristboard [Oct. 30, 2017] has 54 300 subscribers. It posts images once or twice a day, often reposting an image from an Instagram user who used the marketing hashtag #viennanow. We collected the 25 most recent photos posted by ViennaTouristBoard on August 1st, 2017, considering them representative of how Vienna chooses to market itself to the international audience. We both manually and automatically annotated that photo collection in order to compare automatic annotation accuracy. We selected two annotation services: CERTH Image Concept Detection Service (multimedia.iti.gr/mediamixer_images/demonstrator.html [Oct. 30, 2017]) and IBM Watson Visual Recognition Service (visual-recognition-demo.mybluemix.net [Oct. 30, 2017]). In choosing to annotate for the purpose of destination image measurement, it is also important to decide on an appropriate annotation schema. The CERTH service includes results from the Places-205 concept set (places.csail.mit.edu/index.html [Oct. 30, 2017]), developed by MIT CSAIL for scene recognition tasks with 205 scene categories. The service returns for each image (at different levels of granularity: entire video, scene, shot or subshot) a list of concepts with a confidence score (between 0 and 1). IBM Watson’s Visual Recognition Service also returns a set of concepts, termed Classes, with a confidence score. These classes are organised into a type hierarchy and the documentation states only that there are “thousands of possible tags”. Yet there is no canonical list of the classes returned by the service, so we must examine the service outputs ex post facto to assess the quality of the tagging. Examining results from just these two services

highlights a key problem in the image annotation domain: the lack of a common, global annotation schema. The manual classification, performed by the author, aims for an annotation of each image that maximises its value to a tourism provider. Hence this classification focuses on the attributes of the destination that are shown in the media. In a previous work annotating Instagram photos of two destinations (Nixon, 2017), a subset of Beerli's (2004) factors influencing destination image were used in order to define a set of cognitive attributes for the annotation of destination image in visual media. For each of the nine dimensions of Beerli (2004) we extracted the cognitive attributes, leading to an initial list of 53 classes from 7 of the dimensions. In some cases, we were more specific than the Beerli attributes (e.g. from "Flora and Fauna" we extracted the 3 classes Plants and Flowers, Animals and Trees) since we want to distinguish between different attractions at the destinations (e.g. a botanical garden, a zoo or a jungle). Three further classes for visual objects were added during the manual annotation. We thus annotated the Instagram photos three times: once manually (ground truth for the destination image measurement) and twice automatically (using the CERTH and Watson services), where we also determined a mapping between the concepts returned by those services and our reference categorization (of now 56 classes) so that we could compare directly the results.

4 Results

We consider two hypotheses in this evaluation: H1. Online image annotation services can accurately visually annotate destination images; and H2. The visual annotation of destination images is useful for measuring visual destination image. For the first hypothesis, we compare the annotations of the online services with our ground truth annotation, having mapped their annotation responses to our reference categorization of destination image. To measure the accuracy of the annotations, we will use precision and recall calculations for classification tasks, where our task is to classify the image in terms of the (destination image-related) visual concepts it presents to a viewer. The results are shown in Table 1.

Table 1. Accuracy of online image annotation services

Tool	*Precision*	*Recall*	*F-measure*
CERTH	0.60	0.49	0.539
Watson	0.68	0.45	0.542

It may be coincidence that the F-measure of both services result in being similar, however it may be seen that precision of the services (the chance that an annotation made by the service is correct) is higher than the recall (the chance that an annotation in the ground truth is not missed). In the CERTH service, half of the False Negatives were in the Natural Resources category, which may partially be a limitation of the coverage of Places-205 vocabulary (e.g. no annotations for sun/sky, plants and flowers or trees) but also simply having missed some annotations (60% of landscape or water instances were missed). Watson equally had 19 FNs (46% of all FNs) in that category indicating it is particularly challenging for automatic classification. More than half of the CERTH service False Positives were in the classes of Monument and Religion, i.e. classes like tower, basilica or monastery were annotated on images without monuments or religious buildings.

Watson had the least FPs, showing it is more cautious with its annotation of classes to images (NB. for the Watson annotations, we took only classes with a confidence level of 0.5 or more). Overall the Natural Resources category had the least success in annotation (20% accuracy overall) while other visual classes performed well (67% accuracy with Roads, 100% accuracy on Public Transportation and Theme Park, 92% accuracy on Religion, 93% accuracy on Historical Buildings).

Table 2. Measurement of Visual Destination Image

	Natural Resour-ces	General Infra-structure	Tourist Infra-structure	Leisure and recreation	Culture history and art	Politics and economics	Urban enviro-nment
Ground truth	34%	10%	5%	5%	34%	1%	10%
CERTH	12%	14%	0%	5%	58%	0%	11%
IBM Watson	16%	12%	0%	6%	51%	4%	10%

For the second hypothesis, we consider the destination image measurement from the visual annotation. The ground truth annotation is used to determine the ground truth for this measurement. We assume that the comparative frequency of occurrence of the visual classes in the media annotations may be determinant for the visual destination image being presented. To simplify the model, we aggregate the visual class occurrences into the 7 top level categories used by Beerli (2004). Table 2 shows the visual destination image derived from the ground truth and from the annotations of the two image annotation services. The ground truth indicates that destination images posted by the Vienna Tourist Board promote most strongly Vienna's natural resources and its cultural/historical offer. The destination images derived from the automatic services match well on the weaker categories but both vary in the same manner on the two strongest categories, underrepresenting natural resources in imagery and thus over-representing the role of cultural/historical offer in the images, since here visual instances are much more accurately identified.

5 Conclusion

On the basis of this study, based on a small sample of Instagram photos posted by the Vienna Tourism, we may draw some initial conclusions about the usefulness of state-of-the-art online image annotation services for the tourism domain. We focused on the measurement of destination image, a common model for tourism stakeholders to consider how a destination is being presented. While text analysis tools have matured and are being increasingly used by DMOs for this task, multimedia analysis is a "brave new world". Our findings indicate that off-the-shelf solutions are not yet performing as well in the tourism domain as they do in their evaluations reported in the research community, where they are pre-trained on image collections from previously known domains. The coverage of their visual classifiers is a more significant issue than the accuracy of classifiers pre-trained on visual classes, i.e. if the service has been trained on a certain type of visual concept - like a historical building - then the service can generally annotate that concept in images with potentially very high accuracy. However, some types of visual concept are either missing in the training of these services (e.g. characteristics of the natural

environment) or prove more difficult for today's state of the art to detect. As a result, destination images provided by automatic image annotation may skew towards the concepts they do better in detecting, and users need to be aware of the respective capabilities of their chosen image annotation service (e.g. through an initial test of the service with an annotated test dataset which covers all destination image categories) in order to re-balance destination image results accordingly. As future work, we will further annotate destination images and evaluate with an ever larger dataset. This initial experiment has indicated that image annotation services need to be trained specifically for the use case of destination image measurement. We can use our reference categories in creating an appropriate training dataset. Watson's Visual Recognition service can be extended by custom image classifiers, providing us an opportunity to train our own service for the tourism domain. We can benchmark such a tourism-specific image annotation service against the baseline evaluations in this paper to show if a better solution than the state of the art off-the-shelf services can be provided for touristic destination image measurement from multimedia content. As both providers and consumers of destination information use more image and video content, useful services for touristic multimedia annotation will be vital for accurate tourism intelligence in the future.

References

Beerli, A. & Martín, J.D. (2004). Factors Influencing Destination Image. Annals of Tourism Research, 31(3), pp. 657-681.

Fatanti, M.N. & Suyadnya, I.W., (2015). Beyond User Gaze: How Instagram Creates Tourism Destination Brand? Procedia – Social and Behavioral Sciences, 211, pp. 1089-1095.

Hanan, H., Putit, N., (2014). Express marketing of tourism destinations using Instagram in social media networking. In Hospitality and Tourism: Synergizing creativity and innovation in research, pp. 471-474.

Nixon, L., Popova, A. and Önder, I. (2017). "How Instagram influences Visual Destination Image: a case study of Jordan and Costa Rica", ENTER2017 eTourism conference, Rome, Italy

Stepchenkova, S., Li, X., (2012). Chinese outbound tourists' destination image of America: part II. Journal of Travel Research, 6, pp. 687-703.

Stepchenkova, S., Zhan, F. (2013). Visual destination images of Peru: Comparative content analysis of DMO and user-generated photography. Tourism Management. 36. pp. 590-601.

Stepchenkova, S., Li, X., (2014). Destination image: Do top-of-mind associations say it all? Annals of Tourism Research, 45, pp. 46-62

Xiang, Z., Gretzel, U., (2010). Role of social media in online travel information search. Tourism Management, 31, pp.179-188

Acknowledgements

The author thanks Vasileios Mezaris, Evlampios Apostolidis and Konstantinos Apostolidis for access to and support with the CERTH Image Concept Detection Service.