Technology Reports (Special Articles)

Tourist Spot Recommendation System | Tourism Informatics | Neural Network

Special Articles on AI—Expansion of AI Technologies to Diverse Industries and Basic Technologies Supporting AI Applications—

“Generic POI Recommendation”: A Brand-new Deep Learning Approach for Discovering Potential Sightseeing Spots

X-Tech Development Department Hisao Katsumi Wataru Yamada Keiichi Ochiai

Data mining each geographic region for new tourist spots using AI technology can serve as a new solution to social issues such as the overconcentration of tourists in certain spots and having difficulty in attracting tourists to regional cities. In this article, we propose a new idea called generic POIs, alternative sightseeing spots to famous spots. In addition, we propose a method to discover POIs that look similar to those of existing famous sightseeing spots as generic POIs. We also present results of actual mining of Web data for generic POIs and explain our evaluation of the results. The efforts introduced in this article hold great promise as a new step toward realizing a tourism industry that utilizes AI technology going forward.

1. Introduction

The rapid development of means of transportation and the proliferation of mass media have made it possible for us to travel to a variety of places for sightseeing. However, this has led to the issue of overtourism, the overconcentration of tourists in certain tourist sites [1]. Overtourism is a serious issue that adversely affects both tourists and residents living in those sightseeing areas. Three methods are being carried out to effectively address this issue: (1) increase the tourism carrying capacity of existing tourist spots, (2) deconcentrate tourists spatially and temporally at existing tourist spots, and (3) create new tourist spots and induce tourists and visit those destinations [2].

Various efforts are being carried out in the area of (1) increasing tourism carrying capacity, such as building parking facilities at tourist spots. For (2), typical examples of efforts include methods to prevent congestion by presenting real-time crowdedness information in online recommendations of tourist spots [3] and methods to mine for spots that are "well-kept secrets"—sites preferred by local residents more than tourists—by analyzing the attributes of users who post images of tourist attractions on photo sharing websites [4]. Efforts to create new tourist spots and induce tourists and visit them (effort 3) are also important. These efforts include increasing the maximum number of tourist spots in a city by considering spots that had not considered to be tourist sites as tourism resources locations.

As an effort in (3) to create new tourist spots, we at NTT DOCOMO focused on the phenomenon of spots' drawing buzz because they resemble already famous spots and becoming even more popular as a result. For example, Takeda Castle Ruins, famous in Japan, drew media attention because it appears similar to Machu Picchu. As a result, the number of tourists to that spot grew. In this way, if we can find tourism value in spots that had not been considered to date as tourist attractions from the standpoint of their resemblance to famous tourist spots, then it is possible to mine for new tourist spots. At NTT DOCOMO, we call these spots, which can serve as substitutes for already famous spots due to their resemblance, "generic POIs."

In this article, we describe techniques of using image processing AI technology and other technologies to present spots resembling already famous tourist spots as generic POIs. Furthermore, we explain evaluation of our proposed method using images of spots on the Web.

2. Generic POI Extraction Technique

Figure 1 provides an overview of our proposed method. It collects images of candidate spots for mining generic POIs and images of existing famous tourist spots from the Web and calculates the degree of similarity in all combinations of images. Candidate spots that are included in combinations ranked high in similarity are output as generic POIs.

This method and the idea behind it seek to mine for new tourist spots using only images of spots. Compared with general methods for mining tourist spots, our method can mine new tourist spots without depending on the amount of data collected in advance such as word-of-mouth comments and posted images. Furthermore, the method makes it possible to discover new tourism value in spots that even local residents had not recognized as tourist attraction.

3. Generic POI Extraction Technique

This method is composed of the following four processes (Fig. 1).

Process (1): Collect from the Web images of

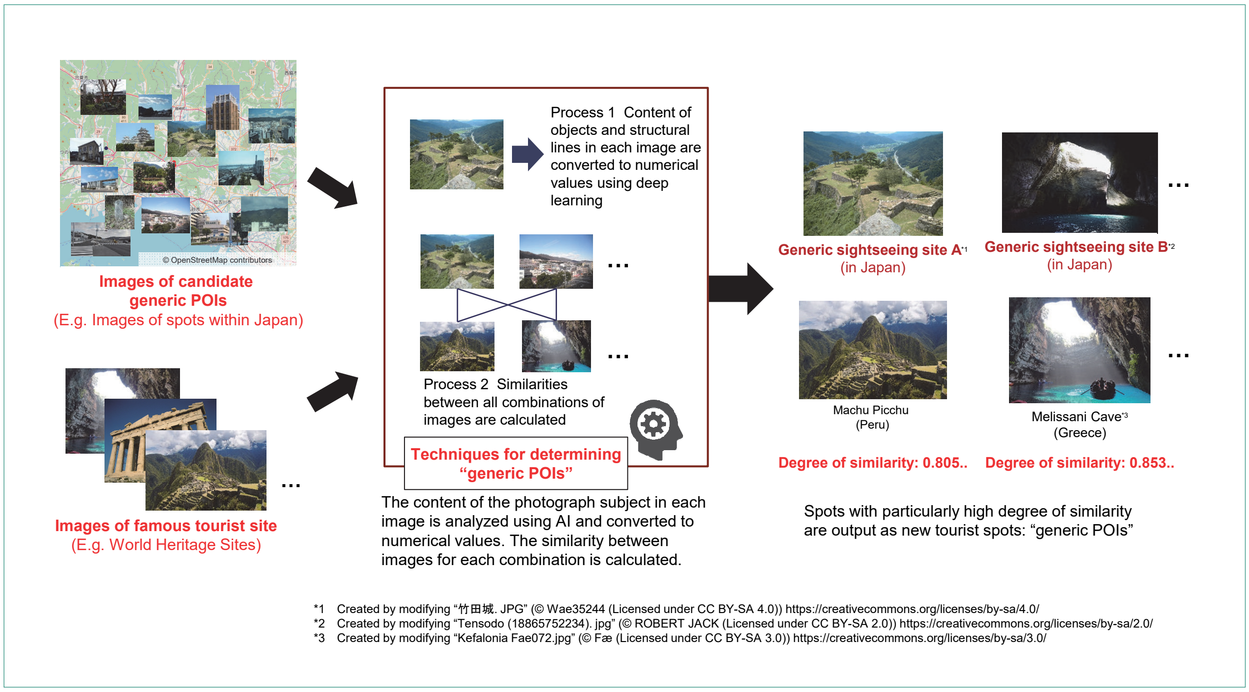


Figure 1 Overview of proposed method

candidate spots taken within an arbitrary distance range.

Process (2): Collect from the Web images of famous spots based on their names.

Process (3): Calculate the similarities between collected images of candidate spots and famous spots for all combinations.

Process (4): Output candidate spots included in the top N combination of similar images as generic POIs.

We assume the use of a variety of images of spots obtained from the Web for Processes (1) and (2). For evaluation of our method, we used images of tourist spots posted on the photo sharing site Flickr*1.

For Process (3), we calculated image similarity for all combinations of candidate spot images and famous spot images. To calculate image similarity, we converted each image of candidate spots and famous spots into feature vectors*2 and used cosine similarity*3 between two feature vectors as the measure of similarity. For feature vectors of tourist spot images, techniques to extract semantic features such as content describing an image and techniques to extract structural features such as structural lines are considered to be useful. For the former, Visual concept is a technique that quantifies the extent to which each image contains each component of scenery such as mountains and rivers in 365 categories [5]. For the latter, computing GIST descriptor is a method to extract feature values such as structural lines and distribution of light and dark areas from an image [6].

Finally, in Process (4), we output candidate site images in top-ranking combinations of images from

---

*1 Flickr: Trademark of Oath Inc. in the U.S.

*2 Feature vector: Representation of patterns and features in data as a vector, which is an array of numerical values, allowing the data features to be handled programmatically.

*3 Cosine similarity: Numerical measure of how close the directions of two vectors are.

all combinations as generic POIs. This method considers as generic POIs to be candidate spots in the top N ranked combinations in image similarity.

4. Evaluation Experiment

To study the validity of our proposed method, we created evaluation data. Furthermore, we mined the evaluation data for generic POIs using the proposed method and evaluated the results by calculating error as the distance between the location of a candidate site and the correct example.

4.1 Creation of Evaluation Data

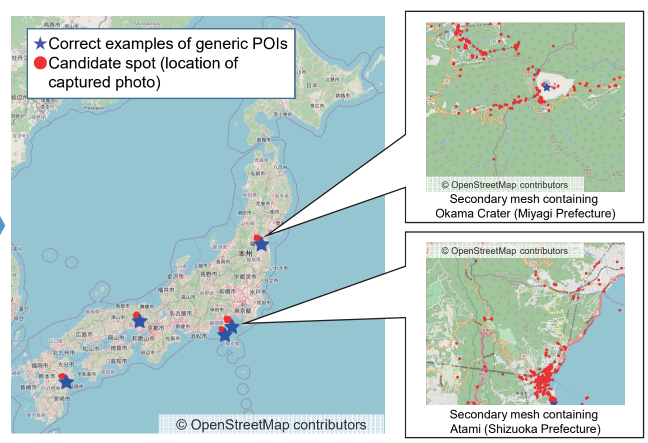
An overview of the method of creating evaluation data is shown in **Figure 2**. First, five famous tourist spots in Japan already known for their resemblance to already famous overseas spots (see table on the left in Fig. 2) were designated as correct examples of generic POIs.

Next using these correct examples, candidate spot images within geographic range of the correct examples of generic POIs were collected from Flickr. Specifically, images taken within a standard region mesh*4 belonging to each of the five correct examples were collected as candidate spot images. For the standard region mesh, we used a secondary mesh that divides the geography of Japan into 10 × 10 km squares. As shown in the right of Fig. 2, the location of a correct generic sightseeing area is marked by ★. Spots where images of candidate spot images were taken were plotted and marked by ●. Using this method, 2,353 candidate spot images were obtained.

To obtain images of the famous spots belonging to the five correct examples, for each of the five spots we entered its English name into the query*5 box in Flickr to get five images each, for a total of 25 images of famous spots. However, the images obtained from search results of famous spots

Famous spots to find resemblance to	Generic POIs
Emerald Lake (New Zealand)	Okama Crater (Miyagi Prefecture)
Amalfi (Italy)	Atami (Shizuoka Prefecture)
Hạ Long Bay (Vietnam)	Dogashima (Shizuoka Prefecture)
Machu Picchu (Peru)	Takeda Castle Ruins (Hyogo Prefecture)
Niagara Falls (U.S./Canada)	Harajiri Falls (Oita Prefecture)

Famous spots and their corresponding generic POIs considered to be correct



Candidate spots gathered based on correct examples

Figure 2 Creation of evaluation data

*4 **Standard region mesh:** Geographic division of Japan into a mesh according to latitudes and longitudes for the purpose of aggregating data in each region.

*5 **Query:** A database query (processing request).

included not only images of the spots themselves but also extraneous images such as people and food, which are considered noise. These noise images were removed manually to prepare the spot images as evaluation data.

Please note that the images collected as described above to create evaluation data were only those licensed by their owners to allow their use for research purposes. Images on Flickr that had the appropriate Creative Commons license*6 type were collected.

4.2 Feature Vectorization of Spot Images

To mine for generic POIs proposed in this article, feature vectorization is carried out for each spot image. The degree of similarity between images is then calculated. For evaluation, we used the following three techniques for feature vectorization and compared the output results obtained.

Technique (1): Visual concepts of scenes

In this method, a neural network*7 is used to determine the extent to which content in each spot image belongs to the 365 categories of scenery, such as mountains, rivers, and the sea. Each spot image is converted into 365-dimensional feature vector. The neural network used here is ResNet18. It carried out training using the Places365 scene recognition dataset [7].

Technique (2): GIST descriptors

Each spot image was converted to a feature vector using GIST. GIST features are extractions of structural features in a spot image, such as the rough distribution of light and dark areas.

Technique (3): Embedding vectors*8 of scene categories

Considering that a neural network extracts features in stages, we used a 512-dimensional vector as the feature vector at the stage before the neural network used in Technique (1) finally computes a 365-feature vector as the visual concept. The feature vector obtained in this manner has higher-level abstraction embedded just before the neural network makes the final inference of 365 scene categories as visual concept (Technique (1)). This processing makes it easier to consider as similar structures that would be classified into different categories under the 365 scene categories.

4.3 Evaluation Results and Discussion

From evaluation data created using the above methods, the proposed method mined for N (N = 10, 20, 30) generic POIs. Photos of generic POIs mined using Technique (3) with N = 30 are shown in **Figure 3** and their locations are shown in **Figure 4**. Within each mesh, we calculated the distance (km) of each mined generic sightseeing site from location of the correct example and obtained an average distance, which we consider as a metric of error. As a generic sightseeing site, the closer in location the mined candidate site is to the correct example, the smaller we consider the error to be. When a spot completely matching the correct example in location is mined, the error becomes 0. However, because spot images collected from the Web includes images taken from different perspectives, images of sites mined from the correct example that are in actuality images of the same

*6 Creative Commons license: Presents creator's intent regarding use of their copyrighted work in accordance with rules set forth by Creative Commons.

*7 Neural network: Mathematical model that imitates the workings of the biological brain, allowing recognition of numerical patterns and making inferences.

*8 Embedding vector: A vector where necessary information is converted from higher-dimensional features to lower-dimensional features.

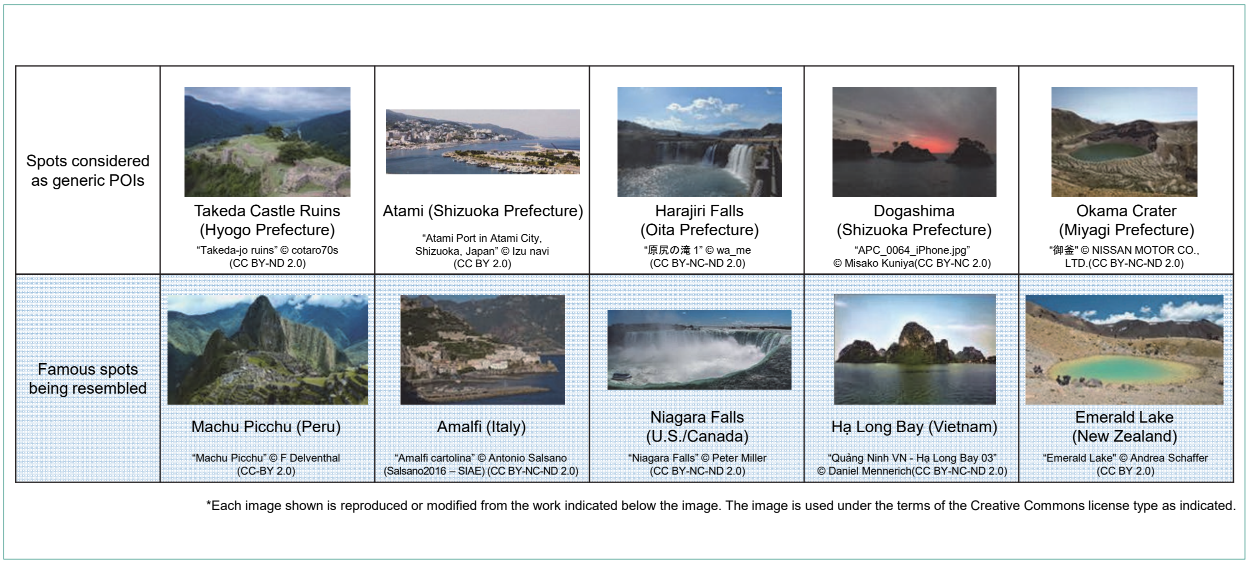


Figure 3 Examples of spots mined as generic POIs using Technique 3 (N = 30)

Figure 4 Calculation of error in location of spots mined as generic POIs using Technique 3 (N = 30)

site may not necessarily have an error of 0. As shown in the right side of Fig. 4, mining the Atami area resulted in several images of spots of Atami that were images of the correct example, a site that resembled Amalfi. These results indicate that these spots were mined as generic POIs of Amalfi. The final results of error evaluation for all the combinations of feature vectorization technique and N candidate images are shown in **Table 1**.

From the results shown in Figures 3 and 4 and Table 1, we see that candidate spots that appear roughly similar to famous spots can be mined as

Table 1 Error calculated for each combination of feature vectorization technique and *N*

	Average error in distance for top 10 spots (km)	Average error in distance for top 20 spots (km)	Average error in distance for top 30 spots (km)
Technique (1) Visual concept	2.18	2.20	2.22
Technique (2) GIST	2.56	2.35	2.45
Technique (3) Embedding vector	0.33	0.27	0.34

generic POIs. As shown in Table 1, Technique (3) produced results that exceeded those of the other techniques. Furthermore, comparing Techniques (1) and (2), in all cases Technique (1) produced better results than Technique (2). These results suggest semantic features, such as content about the subject of a spot photo, are more effective than structural features such as structural lines in an image.

Because our method does not carry out processing of candidate spot images such as clustering*9, it can be applied regardless of the content of input candidate spot images and famous spot images and their quantity. Going forward, we plan to evaluate whether similar results can be obtained when the scale of evaluation data is increased.

5. Conclusion

In this article, we described the idea of data mining for new tourist spots, which we call generic POIs for their resemblance to existing famous spots. We also proposed mining techniques using a neural network, and described evaluation of applying these techniques to tourist spot images posted on the Web. However, as our proposed method was implemented based on five correct locations collected on the Web, we plan to study whether our method can correctly mine a large amount of candidate spot images and famous spot images to find generic POIs, as well as whether mined generic POIs can improve the issue of the overconcentration of tourists and induce them to visit the newly discovered areas.

The proposed ideas and techniques described in this article hold great promise as the first step of efforts that use AI technology to effectively discover new tourism value in spots that had not been considered tourist attractions. They can be expected to be useful for not only the issue of overtourism but also for regional revitalization [8]. In addition, we hope that our proposed ideas and techniques can serve as a step in establishing the new normal*10 in tourism lifestyle, as people accept a new lifestyle that is conscious of preventing the spread of COVID-19. In this new tourism lifestyle, instead of traveling to distant overseas and domestic locations, tourists will see a variety of local spots with a new perspective, becoming cognizant of their heretofore unrealized tourism value and appreciating their hidden allure.

REFERENCES

[1] M. Duignan: "'Overtourism'? Understanding and Managing Urban Tourism Growth beyond Perceptions: Case

*9 **Clustering:** Division of a large amount of data into data groups with similar characteristics.

*10 **New normal:** A state in which a new common sense has irreversibly taken hold as a result of changes in the social environment and circumstances.

Studies," United Nations World Tourism Organisation (UNWTO), pp.34–39, Mar. 2019.

[2] T. Mainil, E. Eijgelaar, J. Klijs, J. Nawijn and P. Peeters: "Research for TRAN Committee-Health tourism in the EU: a general investigation," European Parliament, Directorate General for Internal Policies, 2017.

[3] M. Hidaka, Y. Kanaya, S. Kawanaka, Y. Matsuda, Y. Nakamura, H. Suwa, M. Fujimoto, Y. Arakawa and K. Yasumoto: "n-site Trip Planning Support System Based on Dynamic Information on Tourism Spots," Smart Cities, Vol.3, No.2, pp.212–231, Apr. 2020.

[4] C. Zhuang, Q. Ma, X. Liang and M. Yoshikawa: Anaba An obscure sightseeing spots discovering system," 2014 IEEE International Conference on Multimedia and Expo (ICME), pp.1–6, Sep. 2014.

[5] C. Peters, T. Deselaers, N. Ferro, J. Gonzalo, G. J. F. Jones, M. Kurimo, T. Mandl, A. Penas and V. Petras: "Evaluating Systems for Multilingual and Multimodal Information Access," 9th Workshop of the Cross-Language Evaluation Forum, CLEF 2008, Aarhus, Denmark, Revised Selected Papers, p.527, 2008.

[6] A. Oliva and A. Torralba: "Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope," International journal of computer vision, Vol.42, No.3, pp.145–175, May 2001.

[7] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva and A. Torralba: "Places: A 10 Million Image Database for Scene Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.40, Issue 6, pp.1452–1464, Jul. 2017.

[8] NTT DOCOMO press release: "The first time in Japan! Photo contest to discover scenery in Shikoku that look just like famous overseas spot. In a time when it is difficult to travel far, discover places that give you a sense of traveling abroad! AI judges the similarity of scenery images," Jul. 2021 (In Japanese).