The Champion of Images: Understanding the role of images in the decision-making process of online hotel bookings.

Gijs Overgoor
University of Amsterdam
Bill Rand
North Carolina State University
Willemijn Van Dolen
University of Amsterdam

Cite as:
Overgoor Gijs, Rand Bill, Van Dolen Willemijn (2019), The Champion of Images: Understanding the role of images in the decision-making process of online hotel bookings.. *Proceedings of the European Marketing Academy*, 48th, (9518)

Paper presented at the 48th Annual EMAC Conference, Hamburg, May 24-27, 2019.

The Champion of Images: Understanding the role of images in the decision-making process of online hotel bookings.

We use visual analytics and artificial intelligence to understand the role of images in the decision-making process of consumers booking hotels online. We use deep learning to extract information from hotel images and we apply a visual analytics model to understand the importance of this visual information in the online hotel booking process. We will combine a visual complexity score and a concept score to build an overall image score. This image score could then be used for future firm-generated images to help decide what online travel agencies should show to consumers searching for a hotel. Our method will create an understanding of the role of images in the decision-making process of online hotel bookings. Our framework will allow managers to leverage hotel images to attract consumers quickly and to provide them with visual cues that will facilitate their reservation.

Keywords: Image Analytics, Machine Learning, Online Travel

Track: Methods, Modelling and Marketing Analytics

1. Introduction

Millions of travelers world-wide visit online travel agencies (OTAs) to fullfill their travel needs. OTAs aid the traveler in the decision-making process for the best flight, hotel or vacation package. Working with the hotels and airlines they provide the consumer with a variety of information and travel options. This fast-growing marketplace constituted 600B USD in online travel sales in 2016 and is projected to grow to more than 800B USD in 2020 (Statista 2016). Focusing specifically on online hotel bookings OTAs currently capture 39% of the US online digital booking market (Travel Trends 2017). With such a large consumer demand and stiff competition in the hotel market, if a hotel property or the OTA on which it is hosted wants to be successful, it is essential to attract consumers quickly and to provide them with visual cues that will facilitate their reservation. The image(s) of a hotel have always been part of the hotel listing, even before the Internet, when travel agencies would often have brochures about hotel properties that they used to entice travelers. On many OTA websites, the hotel's image can take up to 33% of the space on the hotel property page, but the importance of this image in the decision-making process has yet to be researched.

Previous research has shown that there are several key attributes that consumers consider for their purchase decisions of hotels including: price, room availability, hotel category, brand, amenities, location, and customer reviews (Kaldis and Kaldis, 2008; Musante, Bojanic, and Zhang, 2009). Researchers have also shown that the Internet enhances the efficiency in search, evaluation, (Christou & Kassianidis 2002; Law 2000, So and Morrison, 2003), which highlights the importance of capturing the consumers' attention. Due to the advanced filtering capabilities, the online traveler is able to quickly evaluate potential candidate hotels through various parameters (Jun, Vogt and Mackay 2010, Varkaris and Neuhofer 2017). Additionally, tracking consumer click, browsing, search, and purchase behavior has shown to be useful in understanding the decision-making process (Ghose, Ipeirotis, and Lui, 2014). Combining these various consumer signals can improve click through rates by optimizing search results (De los Santos and Koulayev 2017). Moreover, presenting advertisement banners that match an online consumer's personality has also been shown to increase conversions (Urban, Liberali, MacDonald, Bordley and Hauser, 2013). Previous research has explored what aspects are important to online travelers and has examined how tracking customer behavior can be leveraged to optimize search results and presentation of information, but very little past research has explored the role of images in travel choice selection. When an online traveler is searching for a hotel, one of the first pieces of information they are presented with in the search results is the hotel image, a thumbnail that appears next to the information presented for the hotel, and given the power of visual information (Khosla, Das Sarma, and Hamid, 2014), it seems clear that this can have an important effect on the customer decision-making process. Therefore, in this paper we will explore the role of the image in the online hotel booking process and we will investigate whether choosing the "right" image can increase click-through rates, volume of bookings and/or booking value.

In OTA terminology, the image that is presented as a thumbnail next to the hotel information and the first image displayed after clicking on the listing, is called the "champion" image. For many OTAs, there are currently no quantitative analytic methods that help determine which image to display in this location. In this research we plan to use machine learning techniques to identify whether it is possible to better understand which images, and specifically which concepts present in those images, are more likely to generate a higher click-through rate, conversion rate, and online sales. We will do this by developing an Image Score for each image that will measure how likely that image is to generate user interest. In addition, to the "champion" image, we could also apply this technique to all the images that are used to represent a property, which would ensure that users see the images that are most likely to have the largest effect first.

2. Framework

In determining the image score we make use of image analytics. We extract several features automatically from the image. These features cover the basic aspect of an image, derived from visual complexity theory and semantic information, using deep neural network structures. In early investigations of images on social media we have found an inverted u-shape relationship between visual complexity and liking behavior. We expect to find similar results between visual complexity and intention to click in an online booking process, but more importantly we hypothesize that the visual complexity information can be helpful for image selection and ordering. Using a novel visual analytics framework, we will extract visual complexity features and classify hotel images using two deep neural network structures: (1) ResNet50 to classify

1000 ImageNet object categories as proposed by He, Zhang, Ren, and Sun (2016) and (2) Places365-ResNet to classify 365 scene categories as proposed by Zhou, Lapedriza, Khosla, Oliva and Torralba (2017).

2.1 Visual Complexity

People's perceptions, preferences and behavior with regards to visual objects, scenes and display are influenced by visual complexity (Machado et al. 2015). Several perceptual features that contribute to visual complexity invoke cognitive and affective responses and they impact preference and liking (Pieters, Wedel and Bartra, 2010; Palumbo, Ogden, Makin, and Bertamini, 2014). Based on these features an image can be perceived to be more or less complex. We automatically extract two categories of complexity (Pieters et al. 2010): Feature Complexity and Design Complexity.

The feature complexity of an image is defined by the detail and variation in the three basic visual features (color, luminance and edges) across an image. The feature complexity is determined by the way an image is taken and it represents unstructured information regarding aesthetics of an image, based on the properties our brain uses to create a structural representation of a scene captured in an image. A photographer might not easily observe some aspects of feature complexity, such as the exact distribution of luminance, when the image is taken, but could be manipulated by a content creator afterwards.. Using our new visual analytics framework we extract color complexity, luminance complexity and edge density from the hotel images. These features measure pixel-level variations in an image and they constitute the feature complexity of an image.

The design complexity of an image captures the visual complexity in terms of the semantic information of the scene in an image. Images with a higher variation in terms of patterns and objects present are more complex. The design complexity focuses on the structured variation in terms of the design whereas the feature complexity focuses on the unstructured variation with regards to the basic image features. The photographer engineers most aspects of design complexity before the shot was taken by composing the shot. We automatically extract the number of unique objects and the visual clutter.

2.2 Image Classification

Recent advances in computer science have provided us the ability to automatically extract conceptual information from a large number of images. This information has shown to be particularly useful in a number of research fields (Khosla, Das Sarma, and Hamid 2014, Li et al. 2014; Mazloom et al. 2016). We use Convolutional Neural Networks to extract conceptual information from the hotel images.

Convolutional Neural Networks (CNNs) are powerful deep learning networks developed primarily for image recognition. CNNs have been successful in identifying objects in images, such as faces, humans and animals. The first CNN, LeNet5, was developed by LeCun et al. (1994). Convolutions are a way of breaking up an image into different areas that focus on processing one particular part of the image. The LeNet5 showed that convolutions are effective at extracting image features in images. Because each convolution is some sort of filter that is applied multiple times to different parts of the image they have only a small set of parameters that need to be estimated to detect similar features in multiple locations in an image. Nowadays, we can use large data sets with labeled images and computer power to learn the parameters in convolutions at a large scale. The CNNs have several types of layers (mathematical manipulations) to extract different types of information from an image. The CNN architecture builds up all kinds of information from the image and combines this for identification. By going back and forth in the model and adjusting weights the CNN can "learn" how to recognize the labeled information in the images of the training set. After a while the model has learned to recognize this same information in new images. For our application, we use two pre-trained CNNs to identify objects and scenes.

The automatic identification of objects in images has received a lot of attention since the start of the ImageNet Large Scale Visual Recognition Challenge (Russakovsky et al. 2014). The challenge evaluates algorithms for object detection and image classification at large scale. They provide a dataset with millions of label images on which these CNNs can be trained. For the identification of objects in hotel images we make use of the pre-trained CNN as proposed by He et al. (2016), the winners of the 2016 Imagenet challenge. The CNN returns a distributional representation of 1000 common objects detected in the image. In other words, for each of the 1000 Imagenet objects it returns a probability score of the particular object being present in the image. The final result is a distributional representation of objects present for every hotel image in our dataset.

For the scene classification we will use a deep neural structure trained on the places database (Zhou et al. 2017). The Places Database consists of 10 million scene photographs, all labeled with scene semantic categories. It comprises a diverse list of types of environment encountered in the world. The deep learning model accurately identifies 365 scene categories depicted in images. Similar to object detection, the pre-trained CNN returns a probability score for the presence of each of the 365 scene categories in the image. The final result is a distributional representation of the presence of scenes for every hotel image in our dataset.

3. Data

Our data consist of consumer searches and the results of those searches for hotels on the website of a global online travel agency. A search starts with a search request. Following the request, which includes several parameters (e.g destination city, travel dates and number of travelers), the website presents the consumer with an ordering of available hotels in the city. Every hotel listing on the search result page consists of the name of hotel, thumbnail or “champion” image, price (with potential discount), number of stars, average reviews. In addition to the standard information that is the provided for every hotel, there is information about deals or specialties unique for that particular hotel or search result. These pieces of special information, such as “free breakfast”, “reserve now pay later”, or “only 1 left at this price”, are often colorfully presented with visual cues such as banners or highlighted text. After obtaining the default set of results, consumers can click on a hotel on that page, continue to the next page of results, or use the sort/filter functionality to refine their results based on hotel characteristics.

At the hotel level we obtain all the characteristics, the champion image as well as the other images of the hotel and the aggregate levels of clicks and bookings for any defined date. At the consumer level we can observe the search criteria, search results, the clicks, view duration and other behavioral data on the website. We are also able to observe the type of device that is used to navigate the page and/or book the hotel.

For our initial exploration, we used 20 hotels in 4 major destinations across the US (Walt Disney World area, Miami, New York and San Francisco). We look at aggregate levels of hotel web visits, meaning the total number of times consumers clicked from the search result page to the hotel page. In future work, based on these results and additional analysis, we will perform experiments on the OTA website and gather more information on the consumer level.

4. Method

One goal of this research is the creation of an image score for each image in the hotel image database. We will combine a visual complexity score and a concept score to build the overall image score. This image score could then be used for future images to help decide which images OTAs should show to consumers searching for a hotel. We train a machine learning model on the historical data. By extracting the visual complexity and concept scores from the images and search results from previously shown images and booking we can learn a model to identify those aspects of images that lead to highest click-through rate (CTR), or any other variable of interest. After training and optimizing our model we can then use this model to construct image scores for new images. Subsequently, we can rank the new images for each hotel based on their image score; choose the champion image based on this information, and then display the images with the highest scores first. Figure 1 illustrates the process of using historic information to train a model and use the image information, hotel- and customer characteristics to score the images, choose a champion image and re-order the other hotel images. We will test this new image scoring method in two ways: First, we can do a holdout test on historical data to see how accurately the image score predicts which hotels users click on. However, that does not really help to improve clicks for a particular hotel and does not control for other differences between hotels. To test the effect of the image score on a particular hotel and as a general model we need to carry out an experiment. This can be done in an experimental setting where for a certain location or a certain type of booking we randomly split incoming searches into two possibilities using an A/B test: (A) control group with no changes, and (B) treatment group where we show the optimal image based on the image score and keep the rest the same. We hypothesize that the hotel with optimal image will see an increase in CTR, sales or booking value over the control group.

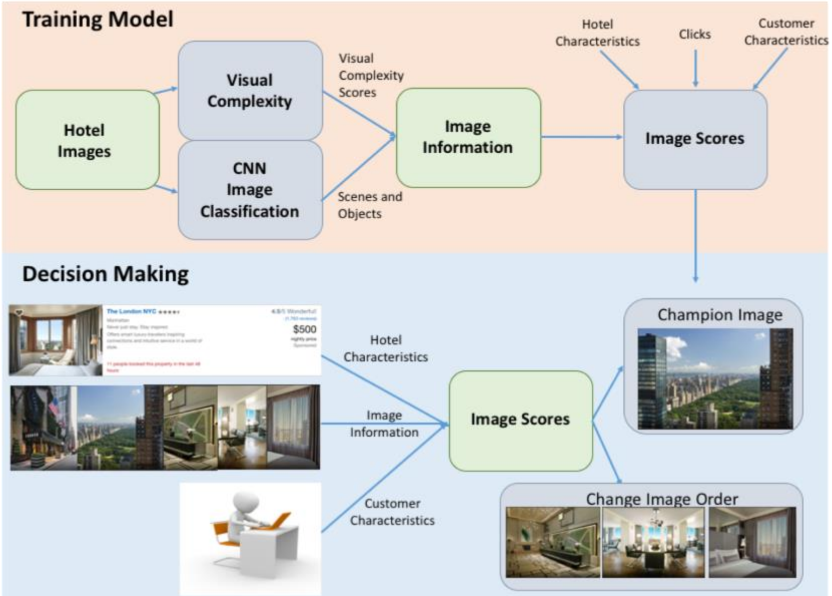

Figure 1. Visualization of the image scoring method.

Findings

We have gathered high-level hotel data such as the CTR, number of views and total bookings a hotel received. For four different locations we examined 20 hotels and gathered the information visible for incoming search on these locations (i.e. price, stars, reviews, etc.) and the "champion image" that is visible on the initial results page. From the champion image we can extract information using image analytics as described in the framework. We will then investigate the relationship between visual complexity (and other visual information) to the overall hotel visits over a defined period. This is, of course, not enough information to use machine learning for the image selection. We will just use these images to do "model-free" exploration of the relationship between image aspects and click-through rate (CTR). We will then apply machine learning after we gather additional data.

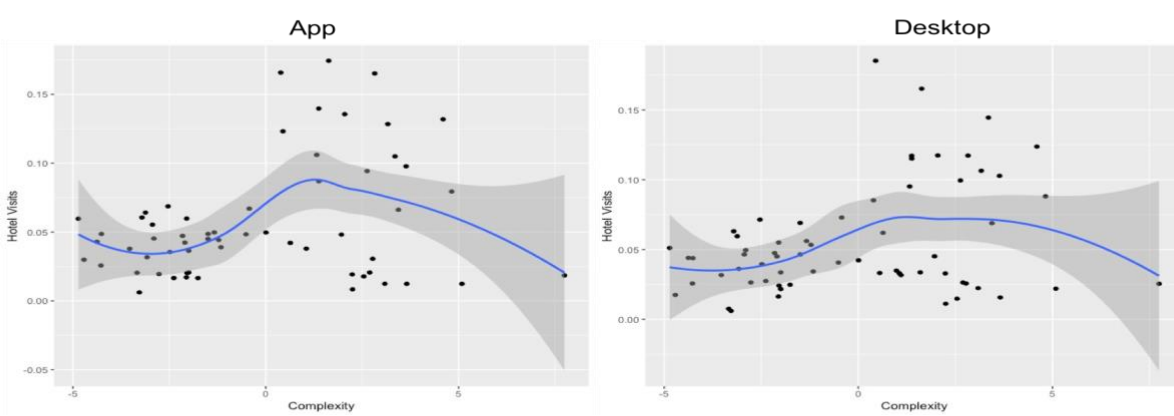

Figure 2. Model-Free Evidence.

Figure 2 contains model-free evidence of the relationship between visual complexity of the images and the CTR. We observe that there is an inverted u-shape relationship between the complexity of an image and the clicks hotels receive from the search result page. The individual level features show more or less the same pattern, except for the unique objects. We did not find a positive effect, or any relationship, for number of objects and the number of online visits the hotel receives. We hypothesize that in the case of online hotel bookings it matters what kind of objects or scene is depicted in the image rather than a number of objects. Therefore, we will investigate this in the next steps. We also do not observe a negative effect for the visual clutter, instead we observe a potential u-shaped relationship for this feature as well. Additionally, there seems to be a more pronounced effect for online visitors using the app(left) than for the online visitors using the desktop (right). This will also be further investigated.

5. Next steps for this paper

Based on our initial explorations into the data, we will design an experiment to investigate the problem on a larger scale and see how we can leverage not just the champion image, but all images and search information to increase clicks and bookings. We will then also apply the deep learning methods to extract the conceptual information. In addition, we explore whether clicks or view duration of a focal user reveal a certain interest for a particular type of booking or hotel image. We can incorporate this behavior into the development of personalized search results. This will enable personalization of images and information being displayed. We will draw from previous work on personalization based on online click behavior (Ghose, Ipeirotis and Lui 2014; De los Santos and Koulayev 2017). These tools give us the ability to gain valuable insight to a large number of images quickly, and to then associate the types of images seen with the likelihood of clicking on the image or making a purchase.

7. References

Christou, E., & Kassianidis, P. (2002). Consumer's perceptions and adoption of online buying for travel products. *Journal of Travel and Tourism Marketing*, *12*(4), 93–107.

De Los Santos, B., & Koulayev, S. (2017). Optimizing click-through in online rankings with endogenous search refinement. *Marketing Science*, *36*(4), 542-564.

Du, R. Y., Hu, Y., & Damangir, S. (2015). Leveraging trends in online searches for product features in market response modeling. *Journal of Marketing*, *79*(1), 29-43.

Ghose A, Ipeirotis PG, Li B (2012). Designing ranking systems for hotels on travel search engines by mining user-generated and crowd-sourced content. *Marketing Science*. 31(3):493–520.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

Jun, S. H., Vogt, C. A., & Mackay, K. J. (2010). Online information search strategies: A focus on flights and accommodations. *Journal of Travel and Tourism Marketing*, *27*(6), 579-595. DOI: 10.1080/10548408.2010.507151

Kaldis, K., & Kaldis, E. (2008). 'Emmantina' and 'Palmyra Beach' Hotels: Distribution for independents hotels. In R. Egger & D. Buhalis (Eds.), *eTourism case stud- ies: Management and marketing issues* (pp. 65–73).

Khosla, Aditya, Atish Das Sarma, and Raffay Hamid (2014). "What makes an image popular?" In: *Proceedings of the 23rd international conference on World wide web*. ACM, pp. 867–876.

Law, R. (2000). Internet in travel and tourism (Part I). *Journal of Travel and Tourism Marketing*, *9*(3), 65–71.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278-2324.

Li, Q., Cai, W., Wang, X., Zhou, Y., Feng, D. D., & Chen, M. (2014, December). Medical image classification with convolutional neural network. In *Control Automation Robotics & Vision (ICARCV), 2014 13th International Conference on* (pp. 844-848). IEEE.

Mazloom, Masoud et al. (2016). "Multimodal Popularity Prediction of Brand- related Social Media Posts". *In: ACM Multimedia Amsterdam, Netherlands*.

Musante, M., Bojanic, D., & Zhang, J. (2009). An eval- uation of hotel website attribute utilization and effectiveness by hotel class. *Journal of Vacation Marketing*, *15*(3), 203–215.

Overgoor, G. et al. (2017). "A Spatio-Temporal Category Representation for Brand Popularity Prediction". In: *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval*.

So, S., & Morrison, A. (2003). Destination marketing organizations' web site users and nonusers: A comparison of actual visits and revisit intentions. *Information Technology & Tourism*, *6*(2), 129–139.

O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. (2014) Imagenet large scale visual recognition challenge. arXiv:1409.0575, 2014.

Urban, G. L., Liberali, G., MacDonald, E., Bordley, R., & Hauser, J. R. (2013). Morphing banner advertising. *Marketing Science*, *33*(1), 27-46.

Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., & Torralba, A. (2017). Places: A 10 million image database for scene recognition. *IEEE transactions on pattern analysis and machine intelligence*.