Applying Visual User Interest Profiles for Recommendation & Personalisation

Jiang Zhou, Rami Albatal, and Cathal Gurrin

Insight Centre for Data Analytics,
Dublin City University
jiang.zhou@dcu.ie
https://www.insight-centre.org

Abstract. We propose that a visual user interest profile can be generated from images associated with an individual. By employing deep learning, we extract a prototype visual user interest profile and use this as a source for subsequent recommendation and personalisation. We demonstrate this technique via a hotel booking system demonstrator, though we note that there are numerous potential applications.

1 Introduction

In this work, we conjecture that images associated with an individual can provide insights into the interests or preferences of that individual, by means of a visual user interest profile (hereafter referred to as the visual profile), which can be utilised to personalise or recommend content to that individual. Given a set of images from an individual, applying deep-learning for semantic content extraction from images generates a set of concepts that form the visual profile. Using this visual profile, it is possible to personalise various forms of information access, such as highlighting multimedia content that a user would potentially like or personalising online services to the particular interests of the user. We believe that it is even possible to develop new (or complimentary) recommendation engines using the visual profile. We demonstrate the visual profile in a real-world information access challenge, that of hotel booking systems. An individual using our prototype hotel booking engine will be presented with visual ranked lists and personalised hotel landing pages. The contribution of this work is the prototype visual profile which imposes no overhead on the user to gather, but can be employed to both recommend content (e.g. hotels to book) and to optimise content (e.g. the hotel landing page).

2 Visual User Interest Modeling

User interest modelling is a topic that has been subject to extensive research in the domain of personalisation and recommendation. While personalisation and recommendation techniques can take many forms, content-based filtering is the approach that best suits our requirements. In content-based filtering, items

are recommended to a user based upon a description of the item and a profile of the user's interests [1]. Content-based recommendation systems are used in many domains, for example, recommending web documents, hotels, restaurants, television programs, and items for sale. Content-based recommendation systems typically support a method for describing the items that may be recommended, a profile of the user that describes the types of items the user likes, and a means of comparing items to the user profile to determine what to recommend.

However, while such recommender systems operate effectively for item-item recommendation, it is our conjecture that a user profile operating at a deeper, more semantic, level than simple item-based user interest profile will capture user interest in more detail and extend recommender and personalisation functionality beyond item-item or faceted recommenders. Hence we introduce the concept of visual user interest modelling which examines media content that are known to be of interest to the user and generates a visual profile of visual concept labels that can be used to subsequently recommend and personalise content to that user. To take a naive example, an individual who regularly views (or captures) images of aircraft, food, architecture, would maintain a visual profile [aircraft, food, architecture] that can be used to highlight/personalise related content when interacting with retrieval systems.

2.1 Visual Feature Extraction

Given a set of images that are associated with an individual, it is necessary to extract the semantic content of the images for the visual profile. Mining the semantic content to extract visual features is an application of content-based image retrieval (CBIR), which has been an active research field for decades [2]. In a naive implementation, low-level image features such as colour, texture, shape, local features or their combination could represent images [3]. Yu et al [4] investigated the weak attributes, a collection of mid-level representations, for large scale image retrieval. Weak attributes are expressive, scalable and suitable for image similarity computation, however we do not consider such approaches to be suitable for generating the visual profile. Firstly, the problem of the semantic gap arises where an individual's interpretation of an image can be different from an individual interpretation. Secondly, the performance of such conventional handcrafted features has plateaued in recent years while higher-level semantic extraction (typcially based on deep learning) has gained favour [5].

Wang et al. [6] proposed a ranking model trained with deep learning methods, which is claimed to be able to distinguish the differences between images within the same category. Given that efficiency is a concern, Krizhevsky and Hinton [7] applied deep autoencoders and transformed images to 28-bit codes, such that images can be fast and accurately retrieved with semantic hashing. Babenko et al. [8] and Wan et al. [9] both proved that pre-trained deep CNNs (Convolutional Neural Networks) for image classification can be re-purposed to image retrieval problem. Babenko used the last three layers before the output layer from CNNs as the image descriptors while Wan chose the last three layers including the output layer.

In our work, we wish to model user visual interest, so our approach is similar to Babenko's and Wan's methods, which also reused a pretrained deep learning network to retrieve and rank hotels images. Our image features are extracted with a CNN [10] where the distribution over classes from the output layer is used as the descriptor for each image. This CNN produces a distribution over 1,000 visual object classes for the visual profile (see Figure 1 for an example of the top 10 object classes for sample images). Because each dimension of the feature vector is actually a class in ImageNet, the descriptor helps to bridge the semantic gap between low-level visual features and high-level human perception.

Instead of training a CNN by ourselves, we employ a pre-trained model on the ImageNet "ILSVRC-2012" dataset from the Caffe framework [5]. Every image is forward passed through the pre-trained network and a distribution over 1,000 object classes from ImageNet is produced. This 1,000 dimension vector is regarded as the descriptor of the image and saved in the visual profile. The visual profile can contain the descriptors of many images.

3 Utilising the Visual Profile

In this work, we represent the prototype visual profile as a collection of feature vectors extracted from a set of images from the individual. Given this set of images, example-based matching between the user profile feature vectors and images in the dataset can be performed. In our case, cosine distance is applied as the similarity metric between pairs of images. The distance is computed with two vectors u and v in an inner product space as Eq. 1. The outcome is bounded between $[0, 1]$. When the images are very similar, the distance approaches to 0, otherwise the value is close to 1.

$$distance = 1 - \frac{u \cdot v}{\|u\|_2 \|v\|_2} \tag{1}$$

Figure 1 shows a simple example of the image matching from our demonstrator system, which assumes a visual profile of one image and a dataset of two images. The first row shows an image (from the user profile) and its top 10 dimensions with highest class probability values in the descriptor. As there is no swimming pool class in the pre-trained model, the model considers the content of the image to be container ship, sea-shore, lake-shore, dam, etc., with meaningful likelihoods. The second row is a similar image from the dataset which also has a swimming pool. As we can observe, classes such as the container ship, dock, boathouse, lakeside are detected in its top 10 classes as well. Not surprisingly, the Cosine distance between these two images is calculated at 0.28 which suggests visual similarity. The third row is a less-related image, for which the top 10 classes has no overlapped with the top 10 of the user profile image. Moreover, the 10 classes are even not semantically close to those from the query image. The distance of 0.98 suggests a high-degree of dissimilarity. By applying this approach on a larger scale, it is possible to select images that are more similar to the visual profile and this knowledge can then be applied to build novel personalised systems and recommendation engines.

Fig. 1: An Image Matching Example using CNNs

4 Applications of the Prototype Visual Profile

It is our conjecture that the use of a visual profile can have many applications when personalising content to an individual. In personal photograph retrieval, the summary given to events or clusters of images can be tailored to the visual profile. In lifelogging, the key images representing an event can be personalised in a retrieval application, or images can be selected from the lifelog that support positive reminiscence. Since image content is modeled as a set of concepts, then the application of this profile can be extended to recommend non-visual content, such as in online stores or social media recommendation. Finally, the application that we find most compelling is to utilise the visual profile to provide a personalised view over, or summary of, visual data. The prototype we chose to develop was in hotel booking, where both the hotel ranked list and the hotel landing pages can be customised to suit the interests of the individual.

4.1 Hotel Booking System: an Example Application

We chose to implement a hotel image recommender system to demonstrate the visual user interest profile in operation. The idea was that a user profile would be augmented by the visual profile, which would be generated from images the user viewed when booking previous hotels, or even from social network postings or online browsing of the user. The visual profile will consist of a weighted list of concepts that occur in images that the user has shown interest in.

The demonstrator both ranked hotels based on similarity to the visual profile, but also personalised every hotel landing page based on the visual profile. In this demonstration, the visual profile is generated by either selecting several hotel images from the interface or uploading images which the user likes. The

recommendation engine will the utilise the visual profile as the searching criteria and return an ranked order hotel list; we compute the minimum distance of N images from the same hotel as Eq. 2 to be the representative distance of that hotel, such that all hotels can be reordered according to the hotel representative distances from 0 to 1 as well.

$$hotel\_distance = \min_{j \in N} \{distance_j\} \tag{2}$$

For each hotel, the images that represent that hotel are also ranked based on the similarity to the visual profile, in effect producing a personalised view of every hotel landing page for each user. Figure 2 shows a screenshot of the demonstrator system. The interface consists of three sections. Across the top of the web page is a set of example image queries to construct the visual profile. These examples cover a wide range of image types (e.g. bedroom images,entertainment systems, pools, dining, etc.). For example, adding swimming pool and pool table in the queries would rank hotels and tailor each hotel page based on these two facilities. The user can also choose to upload any images they wish, which demonstrates the flexibility of the visual profile. The middle section of the interface (Figure 2) is the result frame, which shows the ranked list of hotels or the personalised hotel landing page. Images that are used to construct the visual profile are displayed at the bottom of the window, in which a user can remove or add images to the visual profile before executing a query by selecting 'reorder'.

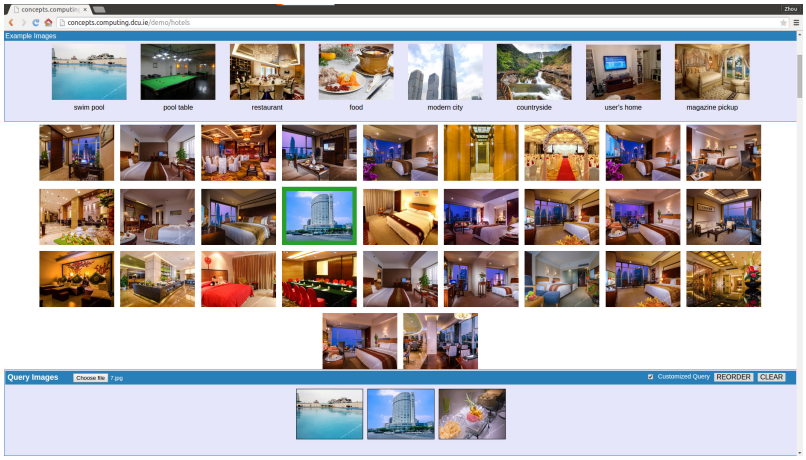

Fig. 2: Applying the Visual Profile for a Prototype Hotel Booking Application. The personalised landing page is shown in this figure.

5 Conclusion & Future Work

In this work we presented a prototype visual user interest profile which attempts to capture the interests of a user by analysing the images that they are known

to like. By employing deep learning, we can extract a visual user interest profile and use this as a source for subsequent recommendation and personalisation. This is novel in that we attempt to capture semantic user interest from visual content, and use it as a means to personalise additional content to the user. We demonstrate this technique via a hotel booking system demonstrator. The next steps in this work are to perform an evaluation of the accuracy of the proposed visual profile in various use cases and applications. We also intend to consider the temporal dimension and explore how time can impact on the weighting in the visual profile and how this impacts on real-world implementations.

Acknowledgments This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under grant number SFI/12/RC/2289.

References

1. Michael J. Pazzani and Daniel Billsus. The adaptive web. chapter Content-based Recommendation Systems, pages 325–341. Springer-Verlag, Berlin, Heidelberg, 2007.
2. Michael S Lew, Nicu Sebe, Chabane Djeraba, and Ramesh Jain. Content-based multimedia information retrieval: State of the art and challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 2(1):1–19, 2006.
3. Thomas Deselaers, Daniel Keysers, and Hermann Ney. Features for image retrieval: an experimental comparison. *Information Retrieval*, 11(2):77–107, 2008.
4. Felix X Yu, Rongrong Ji, Ming-Hen Tsai, Guangnan Ye, and Shih-Fu Chang. Weak attributes for large-scale image retrieval. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2949–2956. IEEE, 2012.
5. Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the ACM International Conference on Multimedia*, pages 675–678. ACM, 2014.
6. Jiang Wang, Yang Song, Thomas Leung, Chuck Rosenberg, Jingbin Wang, James Philbin, Bo Chen, and Ying Wu. Learning fine-grained image similarity with deep ranking. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 1386–1393. IEEE, 2014.
7. Alex Krizhevsky and Geoffrey E Hinton. Using very deep autoencoders for content-based image retrieval. In *ESANN*. Citeseer, 2011.
8. Artem Babenko, Anton Slesarev, Alexandr Chigorin, and Victor Lempitsky. Neural codes for image retrieval. In *Computer Vision–ECCV 2014*, pages 584–599. Springer, 2014.
9. Ji Wan, Dayong Wang, Steven Chu Hong Hoi, Pengcheng Wu, Jianke Zhu, Yongdong Zhang, and Jintao Li. Deep learning for content-based image retrieval: A comprehensive study. In *Proceedings of the ACM International Conference on Multimedia*, pages 157–166. ACM, 2014.
10. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.