Exploring the synergy between knowledge graph and computer vision for personalisation systems

Vincent Lully, Philippe Laublet, Milan Stankovic, Filip Radulovic

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 137 (2018) 175–186

Procedia Computer Science

www.elsevier.com/locate/procedia

SEMANTiCS 2018 – 14th International Conference on Semantic Systems

Exploring the synergy between knowledge graph and computer vision for personalisation systems

Vincent Lully[a,b,*], Philippe Laublet[a], Milan Stankovic[a,b], Filip Radulovic[b]

[a]Sorbonne Université, 28 Rue Serpente, 75006, Paris, France
[b]Sépage, 38 Avenue de l'Opéra, 75002, Paris, France

Abstract

In this paper, we explore the synergy between knowledge graph technologies and computer vision tools for personalisation systems. We propose two image user profiling approaches which map an image to knowledge graph entities representing the interests of a user who appreciates the image. The first one maps an image to entities which correspond to the objects appearing in the image. The second maps to entities which are depicted by visually similar images and which exist in the conceptual scope of the dataset within which further personalisation tasks are conducted. We show the superiority of our second approach against the baseline Google Cloud Vision API (label detection and web entity detection) in terms of accuracy metrics (precision, recall, MRR, nDCG). We also argue the importance of the capacity to create semantically useful profiles as the essence of many knowledge- or semantic-based personalisation systems is the semantic similarity calculation. We then apply our profiling approach in a novel personalisation use case where we seek to select the most appropriate images to display in recommendation banners. Our proposed knowledge-based approach tries to select the images which are the most in line with the semantic user profiles. We hypothesise that this image selection strategy allows to improve the user's perception of the recommended items. We conduct a two-stage user study with a real commercial travel dataset (1,357 package tours in 136 countries and regions depicted by 11,614 images). The results of 32 participants allow us to observe the promising performance of our approach in terms of persuasion, attention, efficiency and affinity.

Keywords: Knowledge graph, computer vision, personalisation, image user profiling, DBpedia, recommeder system, e-tourism.

* Corresponding author. *E-mail address:* vincent.lully@sorbonne-universite.fr

10.1016/j.procs.2018.09.017

1. Introduction

Recent progress in the computer vision (CV) research field has enabled many useful applications such as face detection, content-based image retrieval [1], automatic photo annotation and autonomous cars. Knowledge graphs (KGs) encode human common knowledge in a formally structured form. They are underlying applications like semantic search [2], exploratory search [3], document similarity calculation [4] and question answering [5]. Recent research efforts show several interesting convergence points of these two research fields such as improving object detection with external knowledge graphs [6], scene description with triples [7], knowledge graph completion with visual features [8] and visuo-semantic search [9].

In this paper, we explore the synergy between KG and CV for personalisation systems which has not been sufficiently studied so far. Since several years, KGs have been leveraged to analyse textual data and to conduct personalisation on the web [10-14]. Today, a tremendous amount of multimedia data are available on the web and are being produced continuously. On the Social Web, users publish and share images and videos on different social networks such as Facebook, Twitter, Instagram and Flickr. On e-commerce websites, images are playing an increasingly important role in the discovery of the products and the decision making. Modern websites should be armed with facilities which can understand users' interests through their interactions with multimedia data and adapt the services accordingly in order to provide a better user experience.

The contributions from this paper are two-fold:

- On the one hand, we propose two image user profiling approaches combining KG and CV tools which map an image to knowledge graph entities representing the interests of a user who appreciates the image. We evaluate them qualitatively against a widely used industrial tool Google Cloud Vision API. We show the superiority of one of our approaches in terms of accuracy and argue the importance of the capacity to create semantically useful user profiles as the essence of many knowledge- or semantic-based personalisation systems is the semantic similarity calculation.
- On the other hand, we apply our best-performing profiling approach in a novel personalisation use case that we call "image selection in recommendation banners". Multiple studies and applications have shown the influence of images in the perception of items. The consumer's perception of a hot beverage would be influenced by the color of the plastic vending cup from which it is served [15]. The dating application Tinder shows the best photo as the first photo with the functionality "Smart Photos[1]" and claims to increase matches by 12%. The images hotels choose to display have significant impact on click-through rate[2]. We tackle this new problem with a knowledge-based approach for selecting the images the most in line with users' profiles. We evaluate the approach with a user study and show its promising performance in terms of persuasion, attention, efficiency and affinity.

The rest of the paper is organised as follows. In Section 2, we discuss some related work. In Section 3, we present a general workflow that can be followed by image-based personalisation systems. Section 4 presents the two image user profiling approaches and the qualitative evaluation. In Section 5, we present our approach for image selection in recommendation banners and the user study. Finally, Section 6 concludes the paper.

2. Related work

Given the focus and the contributions of our paper, we present some related work in three directions: approaches combining KG and CV tools, image user profiling, and beyond item recommendations.

2.1. Approaches combining KG and CV

In [6], the authors study the problem of object detection. Compared to existing algorithms which only focus on features within an image, they propose to leverage external knowledge such as knowledge graphs. In [7], the authors

[1] http://blog.gotinder.com/introducing-smart-photos-for-the-most-swipeworthy-you/
[2] https://www.tnooz.com/article/how-image-selection-affects-click-through-rates-on-ota-hotel-listings/

study the problem of structured scene descriptions of images. They combine a statistical semantic model and a visual model. They represent scene descriptions as a set of triples where each triple consists of a pair of visual objects, which appear in the image, and the relationship between them (e.g. man-riding-elephant). In [8], the authors investigate the potential of complementing knowledge graphs with visual features and embeddings learnt from texts. In [9], the authors present a new knowledge graph called “IMGpedia”. It incorporates visual descriptors and similarity relations for the images of the Wikimedia Commons dataset. This dataset enables visuo-semantic queries like “query for images of museums similar to European Catholic cathedrals”.

In our work, different from the use cases treated by previous systems, we combine KG and CV tools for personalisation systems. We found very few work in this area. In [16], the authors study the problem of enhancing point of interest recommendation on location-based social networks using visual contents. The idea is that a user who posts many architecture photos on these social networks is more likely to visit famous landmarks; while a user who posts lots of images about food has more incentive to visit restaurants. They use a convolutional neural network model to extract latent features (4096-dimensional vectors) from images and incorporate them in a machine learning recommendation model.

2.2. Image user profiling

We found several approaches creating user profiles from images. In [17], the authors try to detect demographic attributes of individual users and group types from the photos posted on photo sharing sites. In [18], the authors derive users’ personalities from pictures posted on Instagram. In [19], the authors introduce a picture-based user elicitation and recommendation method for tourism products. The system[3] creates a user profile which consists of 7 traveller types accompanied with a matching degree. A very similar tool is presented in [20] which maps photos to 17 tourist types. In [21], the photos are mapped to several pre-defined categories such as “leisure”, “art” and “culture”.

Different from these existing approaches which map images to demographic characteristics, personalities or pre-defined user types or categories, we propose two approaches which map images to knowledge graph entities which represent users’ interests. This choice has been motivated by some existing work which has proven the advantages of such semantic user profiling in personalisation systems [10-14, 22].

2.3. Beyond item recommendations

For decades, the research efforts in the recommender system field are focused on rating prediction and top-N item selection [23]. More recently, researchers start to be interested in the ways of presenting recommendations and their influence on the perception of recommendations: explanations of the recommendations [24, 25] and page optimization [26]. In our work, we take the initiative to tackle the new image selection in recommendation banners problem with a knowledge-based approach with the aim of enhancing the perception of the recommended items. Our evaluation protocol has been inspired by existing work on explanations [25].

3. Workflow

In Fig.1, we show a general workflow scheme that an image-based personalisation system can follow. The main idea is that through the interactions between users and images, the system creates user profiles by using knowledge graph entities and then makes the advantage of semantic user profiles in different personalisation tasks. In this paper, we only treat the task of image selection in recommendation banners.

[3] https://www.pixmeaway.com/

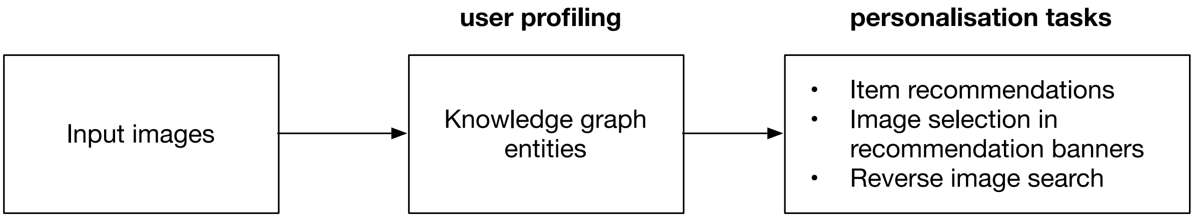

Fig. 1. A general workflow that an image-based personalisation system can follow.

4. Image2Entity: Mapping images to knowledge graph entities

In this section, we present two image user profiling approaches which map an input image to top-n knowledge graph entities. The output entities represent things of interest to a user who appreciates the input image. In this paper, we use the DBpedia, knowing that other similar large-scale knowledge graphs like Wikidata can also be used.

4.1. Approach 1: object detection and entity linking

The first approach consists of mapping an image to entities which correspond to the objects appearing in the image. There are two main steps: object detection and entity linking.

For object detection, we use a computer vision tool named “Inception-V3” [27]. Inception-V3 is a convolutional neural network model trained for the ImageNet Large Visual Recognition Challenge using the data from 2012. The model tries to classify entire images into 1000 classes[4] which are WordNet synsets like “gazelle” and “patio, terrace”.

At the entity linking step, we map the 1000 synsets to corresponding DBpedia entities. 240 of these synsets have known mappings to Wikidata and thus to DBpedia by “owl:sameAs”. For the resting 760 synsets, we made the mappings in a semi-automatic way by referring to the glosses of the synset[5].

We are completely aware that this is a very basic and obvious approach. We still present it because we did not find it in the state of the art. It can be considered as taking the traditionally outputted synsets one step forward and thus enabling the semantic interoperability in knowledge-based systems. However, we can see two obvious shortcomings. On the one hand, it is limited by the number of synsets that can be predicted and it is dependent to the quality of the object detection model. A good range of synsets and a high object detection quality require large-scale training data. On the other hand, it is uncertain that the 1000 synsets are within the conceptual scope of the catalogue (notion detailed just below in 4.2) within which further personalisation tasks are conducted. We propose a second mapping approach to alleviate these shortcomings.

4.2. Approach 2: catalogue-driven visual similarity

The second approach consists of mapping an image to entities which are depicted by visually similar images and exist in the conceptual scope of the catalogue within which further personalisation tasks are conducted.

To compute the visual similarity between images, we rely on the penultimate layer outputted by Inception-V3 which is a 2048-dimensional vector. The similarity between two images is determined by the Euclidean distance between their vectors.

---

[4] http://image-net.org/challenges/LSVRC/2012/browse-synsets
[5] http://image-net.org/archive/gloss.txt

The conceptual scope is a new term that we propose in this work. Given a catalogue of items within which further personalisation tasks are conducted, its conceptual scope consists of all knowledge graph entities which directly appear in the catalogue. The entities can be obtained by two means: direct item linking (for items having corresponding knowledge graph entries like films and artists) and item description linking (for items with rich textual descriptions) as presented in [28]. In case of need, e.g. the number of appearing entities is too small, we may enrich with the entities which are closely related to the appearing ones. The main idea is to map to entities which can contribute to the semantic similarity calculation, in other words, which are useful in further personalisation tasks. Thus, ideally, the conceptual scope should be defined by considering the semantic similarity calculation to be adopted further. For example, we may enrich with the entities by a set of selected object properties [11], or the ones by 1-hop category enrichment [12] or the ones used as dimensions in embeddings [14, 29].

Our second approach requires the following steps:

1. We constitute the conceptual scope of the catalogue by retrieving directly appearing entities (and closed related ones).
2. We retrieve images depicting the entities in the conceptual scope (linked by the property “foaf:depiction”).
3. We compute pairwise visual similarity between the input image and the depicting images with the method explained above.
4. We retain the *n* most similar depicting images and thereafter the entities linked to them.

In Fig.2, we give an example to illustrate the two proposed approaches.

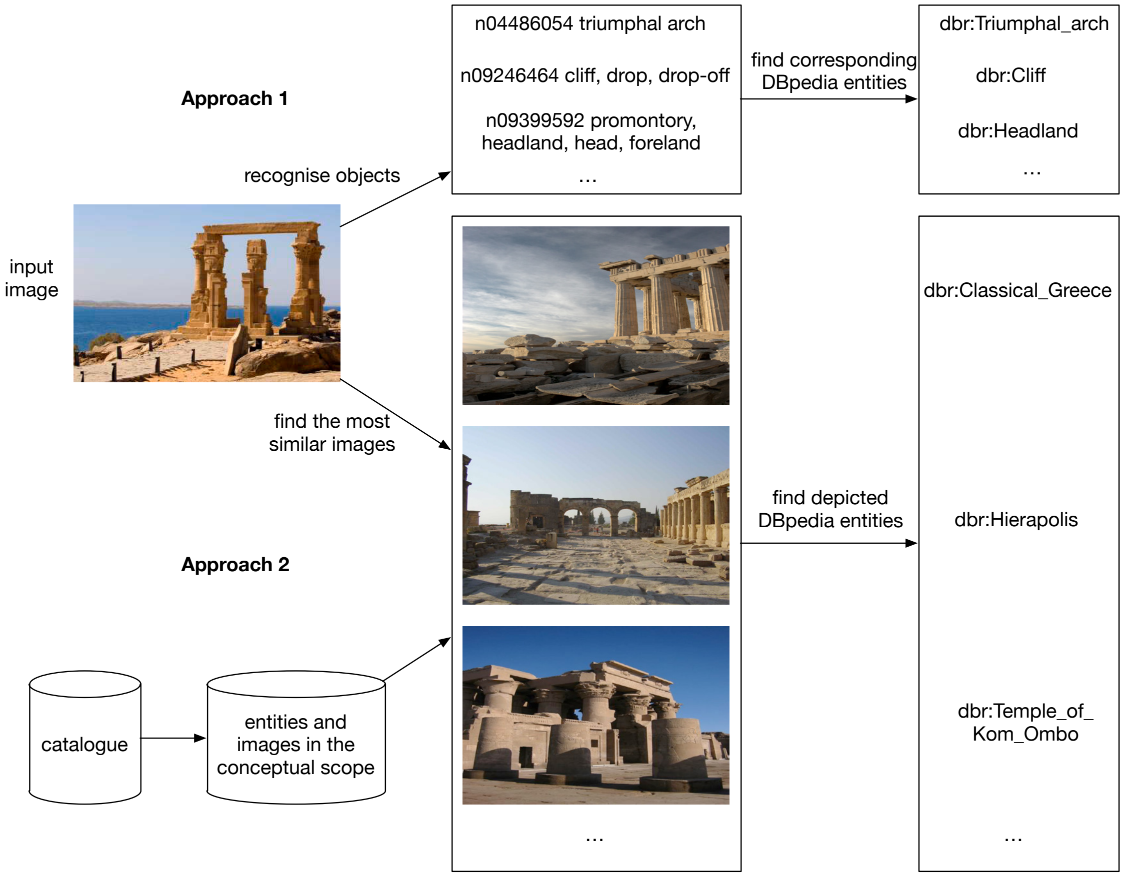

Fig. 2. Example illustrating the two approaches of Image2Entity

4.3. Qualitative evaluation

In this section, we present an experiment to qualitatively check the user profiles created by the two proposed Image2Entity approaches.

4.3.1. Baseline

We failed to find an academic baseline. We compare our approaches with a well-known industrial tool: Google Cloud Vision API[6]. We selected two functionalities: label detection (which identifies objects, locations etc) and web entity detection. As shown in their documentation, the detected labels and entities have known identifiers in Google's knowledge graph. Most of them have equivalents in DBpedia (e.g. "/m/017rgb" corresponds to "dbr:Ferris_wheel"). This makes Google Cloud Vision API more suitable as our baseline and facilitates the comparison of the approaches. For brevity, in the rest of the paper, we may refer to different approaches by using some abbreviations: I2E1 for our first approach, I2E2 for our second approach and GL for the label detection of Google Cloud Vision API and GE for the web entity detection.

4.3.2. Experiment dataset

We use a real and recent commercial catalogue of a popular French travel agency. The catalogue contains 1,357 world-wide package tours which take place in more than 136 countries and regions. The tours are depicted by 11,614 distinct images. Our qualitative check requires human annotators to judge the relevancy of the entities mapped by each approach. It would be very costly to annotate all the images. At this stage of our work, we selected 50 diverse and representative images by using a hierarchical clustering algorithm. For each image, we ran the four approaches (I2E1, I2E2, GL, GE) and retained the top-5 entities. This choice is mainly motivated by the fact that the number of entities outputted by GL and GE is variable but is always at least 5 for the 50 images. We asked three people to manually annotate the relevancy of each entity with regards to each image. To better guide the annotation, they were told to ask themselves the question: *suppose that you like the image, would it be reasonable to determine that you are interested in the entity?* If the answer is positive, the entity is annotated as relevant. Otherwise, it is annotated as irrelevant. The annotations approved by at least 2 annotators are finally retained. The annotators were allowed to check web sources in case they are not familiar with the entities. The created experiment dataset is available online[7].

4.3.3. Process

The created experiment dataset is considered as ground truth. We compare the annotations made by each approach on each image against the ground truth and we calculate scores on multiple metrics.

4.3.4. Metrics

Our measurement is guided by 4 questions and metrics:

1. How relevant are the individual mapped entities? (precision)
2. How many relevant entities are successively mapped? (recall)
3. How early can we find a relevant entity? (mean reciprocal rank)
4. How is the global relevancy of the mapped entities? (normalised discount cumulative gain)

4.3.5. Results

In Table 1, we report the results and provide the scores, the standard deviation and the p-value with respect to the two baselines.

In general, according to all metrics, the performance of the four approaches can be ordered: I2E2 > GL > I2E1 > GE. The differences between the approaches are statistically significant in most of the cases.

---

[6] https://cloud.google.com/vision/
[7] https://github.com/vlully/KG-CV

About baselines. For the baselines, there are two reasons which may explain their less good performance. On the one hand, some mappings have correct labels and wrong entities. For example, for the image (a) in Fig. 3, GL mapped to the entity "/m/02zh30" labelled as "path" whose entity is "dbr:Path_(graph_theory)". On the other hand, some mappings are very generic like "dbr:Building" and "dbr:Travel".

About I2E1. I2E1 achieved a better performance than we expected. Actually we were a little sceptical about the constraints of 1000 entities. It turned out they cover fauna, flora and natural landscape which represent the main spirit the tours of the catalogue.

Table 1. Results of the qualitative evaluation of Image2Entity approaches

	I2E1	I2E2	GL (baseline)	GE (baseline)
precision	0.384	0.576	0.524	0.376
σ	0.279	0.276	0.328	0.282
p-value GL	< 0.01	> 0.1		
p-value GE	> 0.1	< 0.01		
recall	0.226	0.348	0.288	0.223
σ	0.177	0.152	0.138	0.167
p-value GL	< 0.1	< 0.05		
p-value GE	> 0.1	< 0.01		
mean reciprocal rank	0.692	0.886	0.77	0.551
σ	0.407	0.257	0.361	0.406
p-value GL	> 0.1	< 0.05		
p-value GE	< 0.1	< 0.01		
nDCG	0.311	0.478	0.401	0.3
σ	0.245	0.217	0.232	0.24
p-value GL	< 0.05	< 0.1		
p-value GE	> 0.1	< 0.05		

About I2E2. The better performance of I2E2 may be explained by its capacity of capturing the general atmosphere of an image rather than the objects. For example, in the image (b) in Fig. 3 where we can see a train in a rural area, other approaches mapped to entities related to the train object while I2E2 mapped to place entities whose images are visually similar like "dbr:Mokra_Gora". These place entities contain an interesting dimension which goes beyond the mere object entities. In another example image (c) in Fig. 3 where we can see two women wearing traditional clothes, other approaches mapped to object entities like "dbr:Sombrero", I2E2 mapped to ethnic entities like "dbr:Wayuu_people" and "dbr:Toraja". Even though the people in the image are not Wayuu nor Toraja, these entities may still interest the users. I2E2 does not seek to detect the exact entities but what the image globally conveys.

Usefulness of the user profile. An important characteristic of I2E2 is that we limit the spectre of mappable entities to the ones appearing in the catalogue. The results show that this limitation was finally not at the expense of the accuracy. Moreover, we argue the importance of using the entities appearing in the catalogue. Actually the essence of knowledge-based personalisation systems is calculating the semantic similarity between the users and the items [30]. If an entity, however accurate, cannot contribute to the semantic similarity calculation, it is hence useless in such systems. Apart from the main accuracy metrics, we were interested in observing the capacity of different approaches to create user profiles with useful entities. Comparing against only appearing entities would be anecdotal and too favouring I2E2. To make the comparison as fair as possible, besides the 13,109 appearing entities, we included a large amount (2,098,047) of closely related entities by 1-hop category enrichment. We calculate a usefulness of an approach by averaging the usefulness of each entity: 1 if it appears in the catalogue, 0.5 if closely related, 0 otherwise. I2E2 achieved logically the score of 1, I2E1 0.736±0.155, GL 0.732±0.152, GE 0.71±0.21. The outperformance of I2E2 over other approaches is also statistically significant with p-value < 0.01.

Limitations of I2E2. Notwithstanding I2E2 is the best-performing approach, we are aware of two limits. On the one hand, it is dependent to the quality of the entity linking tools. If erroneous entities are extracted and they are depicted by similar images, the quality of the mapping would be deteriorated. On the other hand, it is dependent to the images depicting the entities. The mapping would not be very accurate if the image does not represent well the entity. For example, I2E2 mapped the image (d) in Fig. 3 to “dbr:Oriental_(Morocco)” which is a one of the twelve regions of Morocco. The entity is depicted by an image of a mountainous landscape of Jebel Tamejout which is not representative of the whole region and it is not very reasonable to determine that a user would be interested in the region. For some inclusive geographic entities (like countries) or some abstract entities (like love, history), it is hard to find representative images, this may harm the performance of I2E2.

(a) (b)

(c) (d)

Fig. 3. Images used as examples in the discussion of the evaluation results

5. Image selection in recommendation banners

In recommendation banners, items are often displayed with images. Our hypothesis is that by displaying images more in line with the user’s interests we can improve the user’s perception of recommended items. We propose a knowledge-based approach and evaluate it with a user study.

5.1. Knowledge-based approach

The main idea is to represent a user’s interests (user profile) and an item’s images (image profile) with DBpedia entities. Thus, they are put in the same conceptual space and we can calculate their similarity. Given a user and a recommended item, we select the item’s image whose profile has the highest similarity with the user profile. The image profile is created with the I2E2 approach presented in the previously section. Given the evaluation results, we map each image to the top-3 DBpedia entities. The user profile is created throughout the interaction with the items in the catalogue of the recommender system (as explained in [30]). In this paper, we adopt a naïve approach by linearly aggregating the profiles of the items appreciated in the past. An item profile contains the entities extracted from its description. For all the profiles, we conduct 1-hop category enrichment which has shown to be beneficial in similarity calculation and in recommender systems [11, 13, 28]. We use Jaccard measure to calculate the similarity between different profiles.

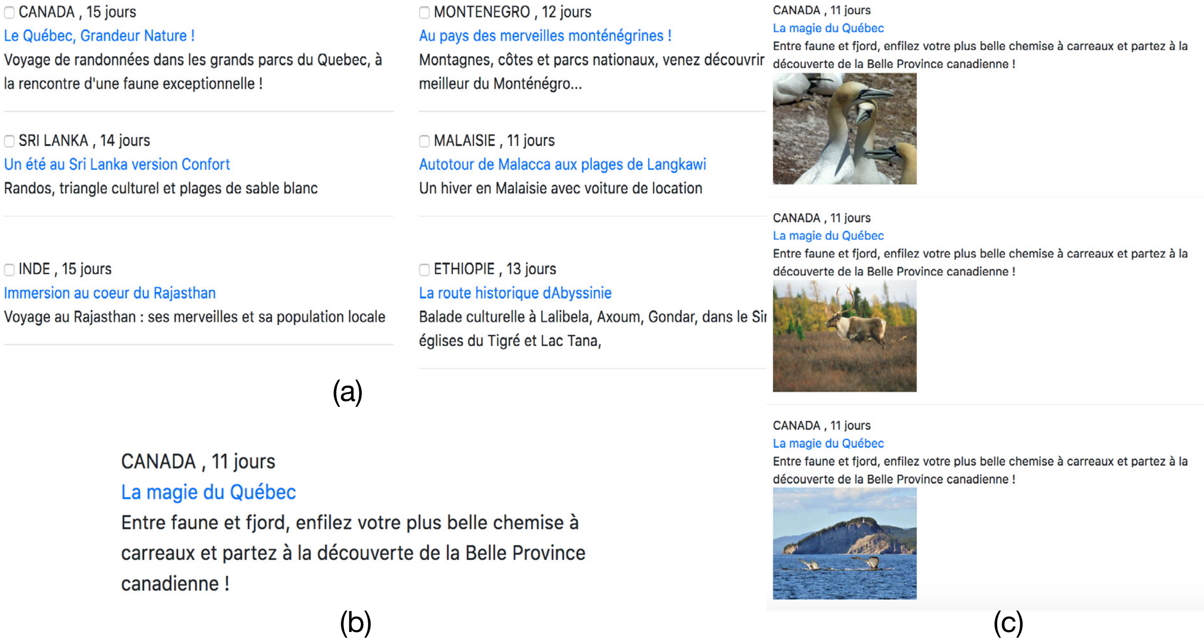

Fig. 4. Evaluation interface of the user study

5.2. Evaluation

We conducted a qualitative use study with the same catalogue as in 4.3. We present successively the baseline, the survey design and the results.

5.2.1. Baseline

We compare the proposed approach (KG) with:

- *Random.* We randomly select an image among available ones.
- *Agent.* We select the image ranked first by the human travel agent who makes the catalogue. This assumes that the travel agent privileges images which are the most attractive in general.

5.2.2. Survey design

Participants put themselves in the scenario of searching for a package tour for their next vacation. They simulate a browsing experience on a web interface where they can visualize the tours (displayed in a random order) where each tour is described by some basic information (image (a) in Fig.4). They can also click on a link to read all the details about the tour. They select one or several tours which appeal to them at first glance. Once submitted, they see a recommendation banner without image (image (b) in Fig.4) and rates on a 5-level Likert scale 3 aspects: persuasion, effectiveness and attention, we call this the first rating stage. After that, they see three recommendation banners with images (image (c) in Fig.4). and rates again persuasion, effectiveness and attention, and also efficiency and affinity, this is the second rating stage. In all banners, the recommended item is the same. All the rated aspects, the statements, the rating stages and metrics are summarized in Table 2. Persuasion, effectiveness and efficiency are taken from the literature of evaluating recommendation explanations [24, 25]. Attention and affinity are two new aspects that we propose in this paper. On the one hand, multiple reports mention the importance of the choice of the images in capturing users' attention [31, 32]. On the other hand, since our selection method privileges images satisfying users' interests, we want to examine how we achieve this goal in a straightforward way.

Table 2. Summary of the related statements in the 2-stage user study

Aspect	Statement	Stage	Metric
Persuasion	I am interested in this recommendation.	1, 2	Rating change
Effectiveness	I have sufficient information to decide whether I click on the recommendation or not.	1, 2	Rating change
Attention	The recommendation banner captures my attention.	1, 2	Rating change
Efficiency	The image helps me decide more rapidly whether to discover more about it or not.	2	Rating
Affinity	The image shows things that I am in affinity with.	2	Rating

5.2.3. Results and discussion

In Fig.5, we show the results of 32 participants (18 women, 14 men, 22 to 32 years old) and we discuss 4 salient points.

Firstly, KG outperforms other approaches on affinity and the difference is statistically significant. This shows that KG is actually capable of selecting an image corresponding to users' interests. Several participants comment that the images selected by KG reflect exactly the trip they imagine during the browsing simulation phase. Secondly, we observe a net increase for all approaches on attention and efficiency. This shows that displaying an image in a recommendation banner can better capture users' attention. Images can help users decide more rapidly whether to discover more about the recommendation or not, as voiced by the participants with the majority positive ratings. Thirdly, the results on effectiveness are not conclusive. For all three approaches, most participants did not find that the images provide much more information to help them decide. The other information displayed in a textual form was sufficient to help them decide whether to click or not. Fourthly, on persuasion, certain participants see their interest in the recommendation decrease after seeing the banners with certain images. Some participants comment that some images are so uninpirational that they even negate the teasing effect of textual descriptions. On this aspect, KG actually enhances the perception of the recommendations while the other two approaches have rather neutral or negative impact. This shows the importance of carefully selecting the images which correspond to users' interests.

6. Conclusion

In this paper, we explored the synergy between knowledge graph and computer vision tools for personalisation systems. We proposed two image user profiling approaches mapping an input image to knowledge graph entities. We showed the superiority of one of our approaches against the baseline Google Cloud Vision API in terms of accuracy and argued the importance of the capacity to create semantically useful profiles. We then took the initiative of tackling the new image selection in recommendation banners problem with a knowledge-based approach which seeks to select the images which are the most in line with the user profiles. We conducted a user study with a real commercial travel catalogue and showed its promising performance in terms of persuasion, attention, efficiency and affinity.

In the future, we envisage diving deeper into the image user profiling approach and studying questions like the impact of the range of mappable entities on the quality of user profile, how to handle inclusive geographic entities and abstract entities whose images may not be representative. We also plan to study the two other personalisation tasks mentioned in Section 3: item recommendation and reverse image search.

Fig. 5. Evaluation results of image selection in recommendation banners

References

[1] Lew, M. S., Sebe, N., Djeraba, C., & Jain, R. (2006). Content-based multimedia information retrieval: State of the art and challenges. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 2(1), 1-19.

[2] Hogan, A., Harth, A., Umbrich, J., Kinsella, S., Polleres, A., & Decker, S. (2011). Searching and browsing linked data with swse: The semantic web search engine. Web semantics: science, services and agents on the world wide web, 9(4), 365-401.

[3] Waitelonis, J., & Sack, H. (2012). Towards exploratory video search using linked data. Multimedia Tools and Applications, 59(2), 645-672.

[4] Paul, C., Rettinger, A., Mogadala, A., Knoblock, C. A., & Szekely, P. (2016). Efficient graph-based document similarity. In International Semantic Web Conference (pp. 334-349). Springer, Cham.

[5] Höffner, K., Walter, S., Marx, E., Usbeck, R., Lehmann, J., & Ngonga Ngomo, A. C. (2017). Survey on challenges of question answering in the semantic web. Semantic Web, 8(6), 895-920.

[6] Fang, Y., Kuan, K., Lin, J., Tan, C., & Chandrasekhar, V. (2017). Object Detection Meets Knowledge Graphs. Ijcai, 1661–1667.

[7] Baier, S., Ma, Y., & Tresp, V. (2017, October). Improving visual relationship detection using semantic modeling of scene descriptions. In International Semantic Web Conference (pp. 53-68). Springer, Cham.

[8] Thoma, S., Rettinger, A., & Both, F. (2017, October). Towards Holistic Concept Representations: Embedding Relational Knowledge, Visual Attributes, and Distributional Word Semantics. In International Semantic Web Conference (pp. 694-710). Springer, Cham.

[9] Ferrada, S., Bustos, B., & Hogan, A. (2017, October). IMGpedia: a linked dataset with content-based analysis of Wikimedia images. In International Semantic Web Conference (pp. 84-93). Springer, Cham.

[10] Di Noia, T., Mirizzi, R., Ostuni, V. C., Romito, D., & Zanker, M. (2012, September). Linked open data to support content-based recommender systems. In Proceedings of the 8th International Conference on Semantic Systems (pp. 1-8). ACM.

[11] Lu, C., Stankovic, M., Radulovic, F., & Laublet, P. (2017). Crowdsourced Affinity: A Matter of Fact or Experience. In European Semantic Web Conference (pp. 554-570). Springer, Cham.

[12] Nguyen, P. T., Tomeo, P., Di Noia, T., & Di Sciascio, E. (2015, October). Content-based recommendations via DBpedia and Freebase: a case study in the music domain. In International Semantic Web Conference (pp. 605-621). Springer, Cham.

[13] Piao, G., & Breslin, J. G. (2016, September). Exploring dynamics and semantics of user interests for user modeling on Twitter for link recommendations. In Proceedings of the 12th International Conference on Semantic Systems (pp. 81-88). ACM.

[14] Ristoski, P., & Paulheim, H. (2016, October). Rdf2vec: Rdf graph embeddings for data mining. In International Semantic Web Conference (pp. 498-514). Springer, Cham.

[15] Piqueras-Fiszman, B., & Spence, C. (2012). The influence of the color of the cup on consumers' perception of a hot beverage. Journal of Sensory Studies, 27(5), 324-331.

[16] Wang, S., Wang, Y., Tang, J., Shu, K., Ranganath, S., & Liu, H. (2017, April). What your images reveal: Exploiting visual contents for point-of-interest recommendation. In Proceedings of the 26th International Conference on World Wide Web (pp. 391-400). International World Wide Web Conferences Steering Committee.

[17] Chen, Y. Y., Cheng, A. J., & Hsu, W. H. (2013). Travel recommendation by mining people attributes and travel group types from community-contributed photos. IEEE Transactions on Multimedia, 15(6), 1283-1295.

[18] Ferwerda, B., Schedl, M., & Tkalcic, M. (2016, January). Using instagram picture features to predict users' personality. In International Conference on Multimedia Modeling (pp. 850-861). Springer, Cham.

[19] Neidhardt, J., Seyfang, L., Schuster, R., & Werthner, H. (2015). A picture-based approach to recommender systems. Information Technology & Tourism, 15(1), 49-69.

[20] Berger, H., Denk, M., Dittenbach, M., Pesenhofer, A., & Merkl, D. (2007, January). Photo-based user profiling for tourism recommender systems. In EC-Web (pp. 46-55).

[21] Linaza, M. T., Agirregoikoa, A., Garcia, A., Torres, J. I., & Aranburu, K. (2011). Image-based travel recommender system for small tourist destinations. In Information and Communication technologies in tourism 2011 (pp. 1-12). Springer, Vienna.

[22] Kapanipathi, P., Jain, P., Venkataramani, C., & Sheth, A. (2014, May). User interests identification on twitter using a hierarchical knowledge base. In European Semantic Web Conference (pp. 99-113). Springer, Cham.

[23] Amatriain, X. (2014, October). The recommender problem revisited. In Proceedings of the 8th ACM Conference on Recommender systems (pp. 397-398). ACM.

[24] Gedikli, F., Jannach, D., & Ge, M. (2014). How should I explain? A comparison of different explanation types for recommender systems. International Journal of Human-Computer Studies, 72(4), 367-382.

[25] Tintarev, N., & Masthoff, J. (2015). Explaining recommendations: Design and evaluation. In Recommender Systems Handbook (pp. 353-382). Springer US.

[26] Lagun, D. (2015). Modeling User Attention and Interaction on the Web (Doctoral dissertation, Emory University).

[27] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2818-2826).

[28] Di Noia, T., & Ostuni, V. C. (2015). Recommender Systems and Linked Open Data. In Reasoning Web International Summer School (pp. 88-113). Springer International.

[29] Cochez, M., Ristoski, P., Ponzetto, S. P., & Paulheim, H. (2017). Global RDF vector space embeddings. In International Semantic Web Conference (pp. 190-207). Springer, Cham.

[30] De Gemmis, M., Lops, P., Musto, C., Narducci, F., & Semeraro, G. (2015). Semantics-aware content-based recommender systems. In Recommender Systems Handbook (pp. 119-159). Springer US.

[31] Lohtia, R., Donthu, N., & Hershberger, E. K. (2003). The impact of content and design elements on banner advertising click-through rates. Journal of advertising Research, 43(4), 410-418.

[32] Pieters, R., & Wedel, M. (2004). Attention capture and transfer in advertising: Brand, pictorial, and text-size effects. Journal of Marketing, 68(2), 36-50.