Journal Homepage : http://dergipark.gov.tr/ejens
EJENS
European Journal of Engineering and Natural Sciences
CODEN: EJENS
DergiPark AKADEMİK

An Image-based Recommender System Based on Image Annotation

Kemal Ozkan[1]*, Zuhal Kurt[2], Erol Seke[3]

[1]*Department of Computer Engineering, Eskişehir Osmangazi University, 26480, Eskişehir, Turkey.*

[2]*Department of Mathematics and Computer Science, Eskişehir Osmangazi University, 26480, Eskişehir, Turkey.*

[3]*Department of Electrical and Electronics Engineering, Eskişehir Osmangazi University, 26480, Eskişehir, Turkey.*

**Corresponding Author email: kozkan@ogu.edu.tr*

Abstract

Recommender system is a software that analyzes available data to make recommendations about various products and services to their users might be interested in. These systems must perform efficient for both users and the e-commerce sites benefiting from such systems. Ensuring proper and reliable recommendations increases user satisfaction that results selling more products and services. Recommendation algorithms can be categorized as collaborative filtering, content-based, and hybrid filtering techniques. Content-based recommendation algorithms are mostly used in text-based systems. However, image-based recommender systems have become increasingly in favour for the content-based recommender systems in recent years.

The process of an image based recommender system is to match a users' images with the most similar images and recommend them. The recommended images are the most likely/similar images that are uploaded and widely acclaimed by the users. The most challenging problem in image-based recommender systems is to match an image with the most similar visual words or classes based on the image's visual content. In this study, we are planning to solve this problem with utilizing bag of words model, which is an effective model in computer vision. Moreover, the visual features are extracted with commonly used descriptors for this model.

Key words

Bag of Words, Descriptor, Similar, Image-based Recommender System.

1. INTRODUCTION

The major aims of recommender systems are helping customers to find products which they seek, and the other one is increasing sales ratios for sellers. Therefore, success of recommender systems is imperative for both users and the e-commerce sites utilizing such systems. Providing accurate and dependable recommendations increases user satisfaction that results selling more products and services. On the other hand, inaccurate recommendations make users to search alternative sites for shopping. Recommender systems can be grouped as collaborative filtering, content-based, and hybrid recommender systems. Collaborative filtering is a common recommender system. In addition to collaborative filtering algorithms, content-based and hybrid recommender systems are increasingly becoming popular due to widespread use of social media. Generally, content-based recommendation algorithms are applied in text-based recommender systems. Recently, image-based recommender systems which utilizes content-based recommendation algorithms, are receiving increasing attention.

Although there are studies focusing on image-based recommender systems, such systems are still challenging research field with some unresolved problems. These systems are based on images uploaded by users. They first determine the most similar users based on the images, which are uploaded to the system by the users before. Then, these systems return the most likely images for the users based on the images liked by their neighbors. The most difficult problem in image-based recommender systems is to match an image with the most similar visual words or classes based on the image's visual content. In this paper, we are planning to solve this problem with using current image annotation techniques, which are effective methods in computer vision. Another challenging problem in image-based recommender systems is the online computation time caused by the large amount of data in real life applications.

An image-based recommender system is proposed that utilizes common image annotation techniques to solve these problems, [1]. This recommendation algorithm works with large collections of touristic images and the structure of this system is based on clustering and classification algorithms. Tourism related images have been annotated geographically, semantically, and visually with utilizing visual features and text data in the system. Then, this system utilizes annotated information to recommend a touristic place where the user can be interested in, [1].

A novel framework called JustClick [3] is a personalized image recommendation system working with large-scale collections of manually annotated Flickr images. The procedure of this system is based on topic networks, that are automatically produced for outlining large-scale collections of manually annotated Flickr images at a semantic level. This system used hyperbolic visualization method to enable interactive navigation and exploration of the topic network. Hence, users can find out the large-scale image collections at the first glance, build up their mental query models interactively and specify their image needs more precisely by selecting the image topics on the topic network directly. Moreover, users can assess the relevance between the recommended images and their real query intentions interactively. An implementation of this system is allowing users to express their time-varying query intentions and to direct the system to more relevant images according to their personal preferences [3].

A personalized recommender system model, that is based on analysis of color features of product images, is proposed in, [2]. This recommendation algorithm is based on colored-feature extraction of images in order to demonstrate impressions related to the human perception of images. This system retrieves and ranks the images corresponding to the desired impressions with utilizing extracted color features. The structure of this model is depend on service-oriented structure in order to promote its flexibility and reuse, that is important when implementing the model into other existing recommender system environments, [2].

A profile based image recommender system is proposed in [4, 5], and this system is able to reach a particular user for retrieval of images from profile based search engines by using smartphone. All users of this system have his/her own profile and semantic signature of each user search is stored individually. This approach improves the performance of search result as it is referring individual profile to know users search intension. The images are re-ranked using keyword expansion to provide better efficiency and effectiveness by using semantic signature for more precise output [4].

The feature extraction techniques are used in order to image annotation, the Color / Scale Invariant Feature Transform (SIFT / C-SIFT) [6,7], Speeded-Up Robust Features [8], Maximally Stable Extremal Regions(MSER) [9] and Local Binary Pattern (LBP) [10] frequently used descriptors, and also new approaches available in the literature. In this paper, we develop an image-based recommender system, and we use novel image annotation approaches to design this system. The proposed system needs to uploaded image from a particular user. These images become query images for our system, then these images are annotated to understand users' preferences. Outputs of the proposed system are recommendation images that are suitable for users' necessity/pleasure. The algorithm of the proposed method and the experimental results will be given in order of the second part and the third part of the study. Finally, the conclusion and future work will be given in section 4.

2. MATERIALS AND METHODS

There are two problems that might affect the overall success of recommender systems. These problems are known as sparsity and accuracy problems. To solve the sparsity problem, we are going to use image annotation techniques to get the image labels, and to construct visual vocabulary. Furthermore, the visual words help the image-based recommender system to collect data about the users, which might improve the accuracy of our recommender system. Contemporary recommender systems utilize GPS data and the labelled data of the collected images to provide recommendations. However, we design our system in such a way so that it will also use unlabeled data to estimate recommendations. At the end, the design of proposed image-based recommender system is based on visual words among images. The current approaches of feature extraction techniques are used to design the proposed image-based recommender system, such as SIFT, SURF, LBP. Overall performance of

the proposed system is evaluated in terms of accuracy and evaluated on a popular dataset. The algorithm of the proposed recommender system is given under below.

- Firstly, a user is prompted to upload the users' images to the proposed system.
- The commonly known feature extraction methods and the bag of Words (BoW) model are used in the proposed system.
- To extract features from the users' images, the popular SIFT, SURF and LBP descriptors are used. Moreover, the proposed system feeds by extracting features from the images in the system.
- In this proposed system, the same feature extraction techniques are applied to users' images, and these images are accepted as test set. Each extracted features from test images are compared to the other images features' in the system. Then top-10 images are recommended to the users.
- The selection opportunity from the recommended top-10 images is given to a user.
- Furthermore, the proposed recommender system is implemented on the Yahoo! shoes dataset, which is generated by Yahoo! Research Lab in 2012.

3. RESULTS AND DISCUSSION

The proposed system is implemented on the challenging Yahoo! Shopping Shoe dataset, [11]. As stated in [11], this dataset provides a new benchmark which contains a diverse collection of types of shoes photos. Object recognition algorithms aim to identify if there is a pair of shoe and the type of shoes (clogs or high heels) appear in a photo automatically. Yahoo! Shopping site is the best place to read user reviews, explore great products and buy online. Dataset is collected from a small subset of product from Yahoo! Shopping to reflect the interesting real-world problem of fine-grained object recognition. This dataset includes a small sample of the Yahoo! Shopping shoe photos. This dataset includes 107 folders, each folder is contained a type and a brand of shoes. Moreover this dataset includes a *.mat file (shoe_annos.mat), which contains a bounding box for each shoe image.

We use shoe annotation file to read images from Yahoo! Shopping Shoe dataset. This file (shoe_annos.mat), contains 4513 image file and a bounding box for each shoe image. For example, the image at classes/lowa_hiking/757550203 640.jpg, has the bounding box: (x1, y1, x2, y2) = (143, 55, 501, 417). We apply boundary box property to all shoe images and to generate a new annotated dataset.

We use SURF, LBP and SIFT descriptor to feature extractions from each image of Yahoo! Shoes Dataset. We select 9 images from each shoe classes to create a training image set, hence the test set includes 21 images from each shoe classes. The descriptors obtained from training images are clustered by k-means clustering method. Clustering centers are considered as visual words that are formed as visual feature vector. To generate image representation, each extracted descriptor is compared to the visual words and associated to the closest word. Finally, image feature histogram vector is obtained. We select randomly 9 images from each class in Yahoo dataset. These 9 images called user's images, we use them as part of test images in our system. The features are extracted from 9 test images with using SIFT, SURF and LBP descriptors. Each extracted features from test images is compared to final image feature histogram vector, then we recommend images which are similar with user's image. The similarity of user images with recommendation images is found with k-nn algorithm. The recommendation accuracy of the proposed recommendation algorithm with utilizing SIFT, Surf, LBP descriptors are respectively %69, %71 and %62. Also, the recommendation accuracy of the proposed system for each class in Yahoo! Shoe dataset is given in Table 1.

For example, we select 10 random images from 7 different heel shoe classes, and then we define these images as user-uploaded images. We recommend 10 shoe images to a particular user, after this recommendation methodology, the user can select a shoe image, which is interested or wanted by this user. An example of the proposed system process is given in Fig. 1, for this example 10 random images from 7 different heel classes are called/named query images, shown Fig. 1. (a). The proposed system recommends images to a particular user with in consideration of query images, hence the recommended images are shown in Fig. 1. (b). It can be concluded in Fig. 1 that only two recommended images are not in the heel shoe classes. Since, Fig. 1 demonstrates that the proposed system can still get good result when the users' images contain various shoes types or brands.

(a) Query Images

(b) Recommendation images

Figure 1. Images are selected from 7 different heel shoe classes as a users'-uploaded images (a) query images, (b) recommended images to a particular user.

Table 1. Recommendation accuracy (%) of the proposed recommender system for each shoe class in Yahoo! Shoe dataset

american eagle sandals	ariat western boots	badgley mischka high heels	betsey johnson platform high heels	birkenstock sandals	clarks sandals	cobian sandals
0,85	0,75	0,9	0,95	0,75	0,8	0,7
1	0,95	0,9	0,8	0,9	0,75	0,8
0,9	0,6	0,95	0,95	0,6	0,8	0,7
corso como flats	creative recreation sneakers	crocs clogs	crocs sandals	crocs slipons	cushe sandals	cushe slipons
0,6	0,6	0,35	0,8	0,2	0,85	0,55
0,75	0,7	0,3	0,9	0,35	0,65	0,65
0,8	0,7	0,4	0,8	0,3	0,7	0,2
ecco laced	ecco slipons	ed hardy sneakers	finn comfort sandals	fitflop sandals	frye laced	grazie sandals
0,3	0,4	0,8	0,9	0,75	0,25	1
0,35	0,4	0,95	0,95	0,9	0,8	0,9
0,05	0,5	0,5	0,75	0,9	0,4	0,9
havaianas sandals	irish setter boots	keen hiking	keen sandals	keen slipons	lacoste slipons	lacoste sneakers
0,75	0,85	0,7	0,7	0,35	0,6	0,35
0,9	0,85	0,45	0,65	0,5	0,65	0,75
0,95	0,6	0,3	0,65	0,25	0,4	0,5
le chameau rain boots	lowa hiking	manolo blahnik high heels	mephisto sandals	mephisto slipons	merrell sandals	merrell slipons
1	0,55	0,9	0,7	0,4	0,85	0,6
1	0,45	0,9	0,9	0,3	0,7	0,5
1	0,25	0,95	0,25	0,2	0,6	0,35
michael kors high heels	nike sandals	oakley sandals	pleaser usa high heels	pleaser usa platform high heels	polo sneakers	puma sneakers
0,8	0,65	0,8	1	0,8	0,5	0,8
0,8	0,8	0,8	1	0,9	0,6	0,7
0,7	0,8	0,95	0,95	0,6	0,4	0,4
quiksilver sandals	ralph lauren sandals	reebok sneakers	reef flip-flops	rider sandals	roxy sandals	sam edelman sandals
0,9	0,8	0,35	0	0,9	0,85	0,8
0,75	0,85	0,55	0,1	0,9	0,9	0,8
0,8	0,75	0,3	0	1	0,8	0,75
sanita clogs	sanuk sandals	sanuk slipons	sebago boat shoes	shiekh sandals	skechers sandals	skechers slipons
0,65	0,95	0,35	0,6	0,95	0,65	0,6
0,5	0,85	0,3	0,5	1	0,75	0,35
0,25	0,7	0,4	0,45	1	0,75	0,15
skechers sneakers	sorel boots	sperry boat shoes	sperry sandals	stuart weitzman high heels	superga sneakers	teva sandals
0,2	0,75	0,6	0,9	0,9	0,7	0,8
0,15	0,65	0,45	0,9	0,858	0,75	0,8
0,45	0,5	0,15	0,7	0,9	0,6	0,65
timberland boots	toms slipons	tory burch sandals	tretorn rain boots	volatile sandals	volcom sandals	yellow box sandals
0,65	0,65	0,8	0,8	0,85	0,75	0,85
0,5	0,8	0,75	0,85	0,8	0,75	0,75
0,4	0,75	0,7	1	0,9	0,85	0,75

4. CONCLUSION

The users are leaving information on the site when on the move, and recommender systems are used this information to recommend products as user's appreciation. Text-based and rating-based recommendation systems attempt to find and recommend items to the users, in contemporary studies. Recently, the number of shopping on the Web is expressed in millions, hence recommender systems are need effective methods to recommend item that user likes or has previously purchased. Hence, an image-based recommender system is examined to provide/propose an effective algorithm for recommender systems. In this paper, a new

recommendation algorithm has been proposed that returns the recommendation images utilizing from the users' uploaded images. After, the users upload images to the proposed system, image annotation techniques are used for extracting features from these images, and then the proposed system recommends images that have same features with the users' images. It is planned to use multi-label image annotation techniques to improve performance of this proposed image-based recommender system.

ACKNOWLEDGMENT

This paper was supported by the Technological and Scientific Research Council of Turkey under its TUBITAK 3001 program with project no: 116E284, entitled 'Developing Image-based Recommender System'. Again, this work was also supported by the Eskisehir Osmangazi University Scientific Research Project Commissions (Grant no.: 2018-1920).

REFERENCES

[1]. K. Pliakos, and C. Kotropoulos, "Building an image annotation and tourism recommender system", *Int. J. Artif. Intell. Tools*, vol. 24 (5), 2015.
[2]. A. Sasa, M. Krisper, Y. Kiyoki, and X. Chen, "A personalized recommender system model using colour-impression-based image retrieval and ranking method", in *Proc. ICIW*, pp.124-130, 2011.
[3]. J. Fan, D.A. Keim, Y. Gao, H. Luo, and Z. Li, "JustClick: Personalized image recommendation via exploratory search from large-scale flickr image collections", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19 (2), pp. 273-288, 2008.
[4]. P. Bhagat, N. Mahakalkar, R. Chaudhari, and A. Gotmare, "A survey paper on profile- based image recommender system for smartphone", *Int. J. of Eng. Res. Tech.(IJERT)*, vol. 3(10), pp. 69-72, 2014.
[5]. L. Cao, J. Luo, A. Gallagher, X. Jin, J. Han, and T.S. Huan, "A worldwide tourism recommendation system based on geotaggedweb photos", in *IEEE ICASSP*, pp. 2274-2277, 2010.
[6]. D. G. Lowe, "Distinctive image features from scale-invariant keypoints", *Int. J. Comput. Vision*, vol.60, pp. 91-110, 2004.
[7]. S. Van de, E.A. Koen, T. Gevers, and G. M.C. Snoek, "Evaluating color descriptors for object and scene recognition", *IEEE T. Pattern Anal.*, vol. 32 (9), pp. 1582-1596, 2010.
[8]. H. Bay, T. Tuytelaars, and L.V. Gool, "SURF: Speeded up robust features", in ECCV, Lecture Notes in Computer Science, Berlin, Germany: Springer, 2006, vol 3951.
[9]. J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions", *Image Vis. Comput.*, vol. 22(10), pp. 761-767, 2004.
[10]. M. Heikkila, M. Pietikainen, and C. Schmid, "Description of interest regions with local binary patterns", *Pattern Recogn.* , vol. 42, pp. 425-436, 2009.
[11]. (2012) Yahoo! Shoes Dataset [Online]. Available: https://webscope.sandbox.yahoo.com/catalog.php?datatype=i.