International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 03 | Mar-2016 www.irjet.net p-ISSN: 2395-0072

One Click Image Search Re-Ranking Based On User Preference

A. Meiappane[1], S. Monesh[2], S. Pradeep Kumar[2], U. Murugan[2]

[1]*Associate Professor, Department of Information Technology, Manakula Vinayagar Institute of Technology Pondicherry*
[2]*Student, Department of Information Technology, Manakula Vinayagar Institute of Technology, Pondicherry*

Abstract - *Web-scale image search engines (For e.g. Google Image Search, Bing Image Search, Pinterest) mostly rely on surrounding text features. It is difficult for them to predict user's intention only with the query they are giving and this leads to ambiguous and noisy search results from the search engines which are far from satisfactory. In this paper, we propose a novel Internet image search approach. Which requires the user only to click on one query image with the minimum effort and images from the database retrieved by text-based search are re-ranked based on both visual and textual content. Our goal is to capture the user's search intention from this one-click query image in four steps as follows. (1) The query image which the users search for is categorized into one of the predefined adaptive weight categories, which reflect user's search intention at some level. (2) Based on the visual content of the searching image selected by the user and through image clustering mechanism, the query keywords are expanded to capture user intention by one click. (3) Expanded keywords are used to enlarge the image pool to contain more relevant images in which the user search for. (4) Expanded keywords from the above stage are also used to expand the query image to numerous positive visual examples from which new query specific visual and textual similarity metrics are learned to further improvise the results of content-based image re-ranking for improvised results. All these steps are automatic without extra effort from the user.*

Key Words: **Image search, Intention, User preference, Image re-ranking, Adaptive similarity, Keyword expansion**

1.INTRODUCTION

Many commercial image search engines in the internet use only keywords as queries. Users type query keywords as input in the hope of finding a certain type of images they search for. The search engine returns images in thousands that are ranked by the keywords extracted from the surrounding text. It is well known that text-based image search process suffers a lot from the ambiguity of query keywords. The keywords provided by users tend to be short and mostly not commonly known. For example, the average query length of the top 2, 000 queries of Picsearch is 1.369 words, and 95% of them contain only one or three words [1]. They cannot describe the content of images accurately and perfectly. The search results are noisy and ambiguous consist of images with quite different semantic meanings. Fig1 shows the top ranked images that are ranked from Bing image search using "Jaguar" as query. They belong to different categories, such as "Blue Jaguar car", "Black Jaguar car ", "Jaguar logo", and "Jaguar animal", due to the ambiguity of the word "Jaguar". The ambiguity issue occurs for so many reasons.

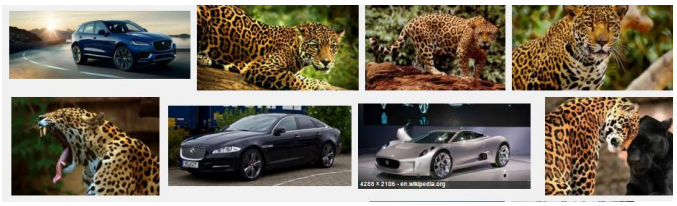

Fig 1. Top ranked images returned from Bing image search using "Jaguar" as query.

First, the query keywords that the user searching for, meanings may be richer than users' expectations. Consider this, the meanings of the word "Jaguar" includes Jaguar animal and Jaguar car and Jaguar logo. Second, the user may not have enough knowledge about the textual description of target images he/she searching for. The most importantly, in many scenarios, it is difficult for users to explain the visual content of queried images using keywords accurately. In order to solve the ambiguity issues, additional information has to be used. One way is text-based keyword expansion , that makes the textual description of the query more detailed. Existing linguistically-related methods find either synonyms and other linguistic-related words from thesaurus state, or finds words as frequently co-occurring with the query keywords. However, the interaction between the user and the system has to be as simple as possible. The minimum criteria is that a One- Click. In this paper, we propose a kind of novel Internet image search approach. It just requires the user to give only a click on a query image and images from a dataset or a pool is retrieved by text-based search are re-ranked based on their visual and textual similarities to the query image searching for. The users will tolerate one-click interaction which has been used by many famous text-based search engines. For example, in Google it requires a user to select a suggested textual query expansion by one-click to get additional results as output. The problem solved in this paper is how to capture user intention from this one-click query image.

2. Literature Survey

2.1 Image Search and Visual Expansion

Many Internet scale image search methods [11]–[17] are text-based and are limited by the fact that query keywords cannot describe image content accurately. Content-based image retrieval uses visual features to evaluate image similarity. Many visual features [5]–[9] were developed for image search in recent years. Some were global features such as GIST [5] and HOG [6]. Some quantized local features, such as SIFT [13], into visual words, and represented images as bags-of-visual- words (BoV) [8]. In order to preserve the geometry of visual words, spatial information was encoded into the BoV model in multiple ways. For example, Zhang et al. [9] proposed geometry-preserving visual phases which captured the local and long-range spatial layouts of visual words.

One of the major challenges of content-based image retrieval is to learn the visual similarities which well reflect the semantic relevance of images. Image similarities can be learned from a large training set where the relevance of pairs of images is known [10]. Deng et al. [11] learned visual similarities from a hierarchical structure defined on semantic attributes of training images. Since web images are highly diversified, defining a set of attributes with hierarchical relationships for them is challenging.In general, learning a universal visual similarity metric for generic images is still an open problem to be solved.

Some visual features may be more effective for certain query images than others. In order to make the visual similarity metrics more specific to the query, relevance feedback [12]–[16] was widely used to expand visual examples. The user was asked to select multiple relevant and irrelevant image examples from the image pool. A query-specific similarity metric was learned from the selected examples. For example, in [12]–[14], [16], [17], discriminative models were learned from the examples labeled by users using support vector machines or boosting, and classified the relevant and irrelevant images. In [21] the weights of combining different types of features were adjusted according to users' feedback. Since the number of user-labeled images is small for supervised learning methods, Huang et al. [15] proposed probabilistic hypergraph ranking under the semi-supervised learning framework. It utilized both labeled and un- labeled images in the learning procedure. Relevance feedback required more users' effort. For a web-scale commercial system users' feedback has to be limited to the minimum, such as one-click feedback.

In order to reduce users' burden, pseudo relevance feedback [18], [19] expanded the query image by taking the top *N* images visually most similar to the query image as positive examples. However, due to the well- known semantic gap, the top *N* images may not be all semantically-consistent with the query image. This may reduce the performance of pseudo relevance feedback. Chum et al. [8] used RANSAC to verify the spatial configurations of local visual features and to purify the expanded image examples. However, it was only applicable to object retrieval. It required users to draw the image region of the object to be retrieved and assumed that relevant images contained the same object. Under the framework of pseudo relevance feedback, Ah- Pine et proposed trans-media similarities which combined both textual and visual features proposed the query-relative classifiers, which combined visual and textual information, to re-rank images retrieved by an initial text-only search. However, since users were not required to select query images, the users' intention could not be accurately captured when the semantic meanings of the query keywords had large diversity.

We conducted the first study that combines text and image content for image search directly on the Internet, where simple visual features and clustering algorithms were used to demostrate the great potentional of such an approach. Following our intent image search work in [1] and [2], a visual query suggestion method is developed. Its difference from [1] and [2] is that instead of asking the user to click on a query image for re-ranking, the system asks users to click on a list of keyword-image pairs generated off-line using a dataset from Flickr and search images on the web based on the selected keyword. The problem with this approach is that on one hand the dataset from Flickr is too small compared with the entire Internet thus cannot cover the unlimited possibility of Internet images and on the other hand, the keyword-image suggestions for any input query are generated from the millions of images of the whole dataset, thus are expensive to compute and may produce a large number of unrelated keyword- image pairs.

Besides visual query expansion, some approaches used concept-based query expansions through map- ping textual query keywords or visual query examples to high-level semantic concepts. They needed a pre-defined concept lexicons whose detectors were off-line learned from fixed training sets. These approach were suitable for closed databases but not for web-based image search, since the limited number of concepts cannot cover the numerous images on the Internet. The idea of learning example specific visual similarity metric was explored in previous work. However, they required training a specific visual similarity for every example in the image pool, which is assumed to be fixed. This is impractical in our application where the image pool returned by text based search constantly changes for different query keywords. Moreover, text information, which can significantly improve visual similarity learning, was not considered in previous work.

2.2 KEYWORD EXPANSION

In our approach, keyword expansion is used to expand the retrieved image pool and to expand positive examples. Keyword expansion was mainly used in document retrieval. Thesaurus based methods expanded query keywords with their linguistically related words such as synonyms and hypernyms. Corpus-based methods, such as well known

term clustering and Latent Semantic Indexing, measured the similarity of words based on their co-occurrences in documents. Words most similar to the query keywords were chosen as textual query expansion. Some image search engines have the feature of expanded keywords suggestion. They mostly use surrounding text.

Some algorithms generated tag suggestions or annotations based on visual content for input images. Their goal is not to improve the performance of image re- ranking. Although they can be viewed as options of key- word expansions, some difficulties prevent them from being directly applied to our problem. Most of them assumed fixed keyword sets, which are hard to obtain for image re-ranking in the open and dynamic web environment. Some annotation methods required supervised training, which is also difficult for our problem. Different than image annotation, our method provides extra image clusters during the procedure of keyword expansions, and such image clusters can be used as visual expansions to further improve the performance of image re-ranking.

2. EXISTING SYSTEM

2.1 CLICK-BASED RELEVANCE FEEDBACK

Inspired by the retrieval approach Pseudo-Relevance Feedback (PRF), we propose a novel re-ranking algorithm, called Click-Based Relevance Feedback (CBRF), which not only sufficiently leverages click-through data but also adequately exploits the interactions between multiple modalities. We begin this section by an overview of our proposed CBRF, elaborate multi-modality fusion with simple MKL embedded in CBRF, and give the algorithm details of CBRF. The framework overviews of PRF and CBRF are illustrated on the left and right plot of Fig 1 . As Fig 1 shows, PRF treats the top (bottom) ranked images from the initial ranked list as pseudo-positive (pseudo-negative) data. With these training data, generally, PRF leverages an individual modality to perform re-ranking, which means the number of SVMs is equal to 1, i.e., n = 1.When working with multiple modalities, PRF uses multiple SVMs correspondingly, and then it simply fuses the outputs to and from different SVMs in a linear way and as the final re-ranking scores. In contrast, the Fig 2 shows, CBRF treats clicked images as pseudo-positive data and randomly selects images from other queries as pseudo-negative data.

Overview

Similar to pseudo-relevance feedback, the proposed click-based relevance feedback treats image search re-ranking as a binary classification problem, where positive data consist of relevant images in the collection and negative data the irrelevant ones. The main idea of CBRF is to leverage click-through data of the related query as pseudo-positive data and apply a multiple kernel learning (MKL) algorithm to learn suitable query-dependent fusion weights for multiple modalities.

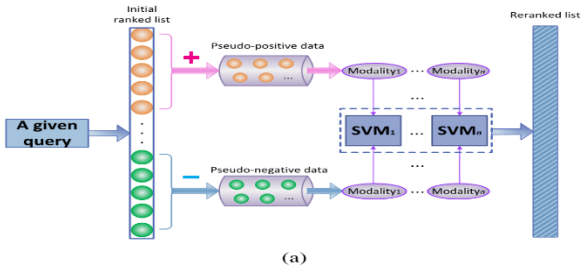

Fig 1 Framework Overview of PRF

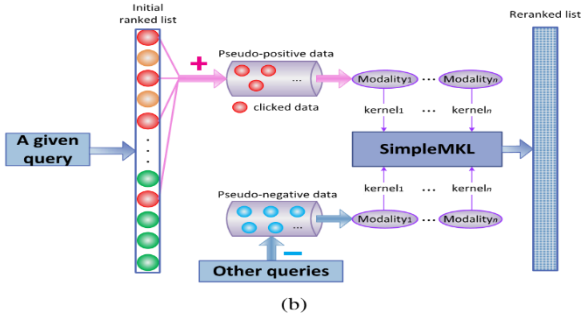

Fig 2 Framework Overview of CBRF.

For a given query, as Fig 2 demonstrates, when collecting training data, CBRF first chooses the clicked images (denoted by red circles in Fig 2) as pseudo-positive data, and then randomly selects images from other queries (denoted by blue circles in Fig 2) in the dataset as pseudo-negative data. With these training data, multiple visual features extracted offline are added into a simpleMKL classifier, which aims to sufficiently explore the influence of different modalities for the given query. As we can see from Fig 2, before multi-modality fusion, each modality is assigned a specific kernel according to its data distribution. Then, through the process of multiple kernel learning, the simpleMKL classifier predicts the fusion weights of multiple modalities adaptively and query-dependently. With an appropriate combination of multiple modalities, the classifier outputs the re-ranking scores of images from the initial ranked list, which are equal to the posterior probabilities of images classified as positive data. This whole re-ranking process can be considered a partially supervised learning from the viewpoint of machine learning.

3. PROBLEM DEFINITION

Experiments conducted on a real world dataset not only demonstrate the usefulness of click-through data, which can be viewed as the footprints of user behavior, in understanding user intention, but also verify the importance of query-dependent fusion weights for multiple modalities. Moreover, significant performance improvement using our proposed re-ranking approach is observed in most query

types in our dataset compared with other re-ranking approaches, which validates the effectiveness and superiority of our approach. In this paper, we only take image search relevance into consideration, though image diversity is another important factor in search performance. In future work, we will focus on enhancing the diversity of re-ranked images by duplication detection or other such method.

4. PROPOSED SYSTEM

Keyword-based search has been the most popular search paradigm in today's search market. Despite simplicity and efficiency, the performance of keyword-based search is far from satisfying. Investigation has indicated its poor user experience - on Google search, for 52% of 20,000 queries, searchers did not find any relevant results [1]. This is due to two reasons. Toy example for non-personalized (top) and personalized (bottom) search results for the query "jaguar"."IR" has the interpretation of both information retrieval and infra-red. Users may have different intentions for the same query, e.g., searching for "jaguar" by a car fan has a completely different meaning from searching by an animal specialist. One solution to address these problems is *personalized search*, where user-specific information is considered to distinguish the exact intentions of the user queries and re-rank the list results. Given the large and growing importance of search engines, personalized search has the potential to significantly improve searching experience.

4.1 OVERVIEW OF THE PROJECT

Increasingly developed social sharing websites, like Flickr and You tube, allow users to create, share, annotate and comment Medias. The large-scale user generated meta-data not only facilitate users in sharing and organizing multimedia content, but provide useful information to improve media retrieval and management. Personalized search serves as one of such examples where the web search experience is improved by generating the returned list according to the modified user search intents. In this paper, we exploit the social annotations and propose a novel framework simultaneously considering the user and query relevance to learn to personalized image search. The basic premise is to embed the user preference and query-related search intent into user-specific topic spaces. Since the users' original annotation is too sparse for topic modeling, we need to enrich users' annotation pool before user specific topic spaces construction.

The proposed framework contains two components:

- ***A Ranking based Multi-correlation Tensor Factorization model*** is proposed to perform annotation prediction, which is considered as users' potential annotations for the images;
- We introduce ***User-specific Topic Modeling*** to map the query relevance and user preference into the same user-specific topic space.
- For performance evaluation, two resources involved with users' social activities are employed. Experiments on a large scale dataset demonstrate the effectiveness of the proposed method.

4.2 MODULES

- Image Search
- Query Categorization
- Visual Query Expansion
- Images Retrieved by Expanded Keywords

Image Search

In this module, Many Internet scale image search methods are text-based and are limited by the fact that query keywords cannot describe image content accurately. Content-based image retrieval uses visual features to evaluate image similarity.

One of the major challenges of content-based image retrieval is to learn the visual similarities which well reflect the semantic relevance of images. Image similarities can be learned from a large training set where the relevance of pairs of images.

Query Categorization

In this module, the query categories we considered are: General Object, Object with Simple Background, Scenery Images, Portrait, and People. We use 500 manually labeled images, 100 for each category, to train a C4.5 decision tree for query categorization. The features we used for query categorization are: existence of faces, the number of faces in the image, the percentage of the image frame taken up by the face region, the coordinate of the face center relative to the center of the image,

Visual Query Expansion

In this module, the goal of visual query expansion is to obtain multiple positive example images to learn a visual similarity metric which is more robust and more specific to the query image. The query keyword is "Paris" and the query image is an image of "eiffel tower". The image re-ranking result based on visual similarities without visual expansion. And there are many irrelevant images among the top-ranked images. This is because the visual similarity metric learned from one query example image is not robust enough. By adding more positive examples to learn a more robust similarity metric, such irrelevant images can be filtered out. In a traditional way, adding additional positive examples was typically done through relevance feedback, which required more users' labeling burden. We aim at developing an image re-ranking method which only requires

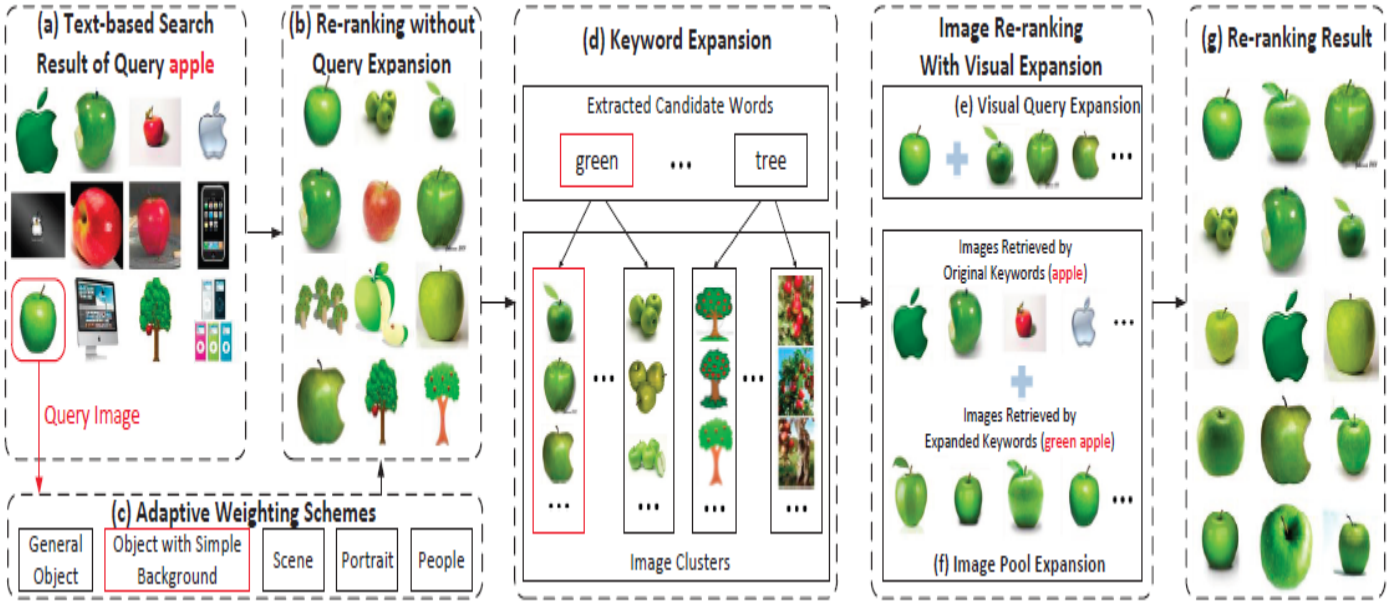

Fig 3 Architecture Of One Click Image Search

one-click on the query image and thus positive examples have to be obtained automatically.

Images Retrieved by Expanded Keywords

In this module, considering efficiency, image search engines, such as Bing image search, only re-rank the top *N* images of the text-based image search result. If the query keywords do not capture the user's search intention accurately, there are only a small number of relevant images with the same semantic meanings as the query image in the image pool. Visual query expansion and combining it with the query specific visual similarity metric can further improve the performance of image reranking.

5. CONCLUSION

How to effectively utilize the rich user metadata in the social sharing websites for personalized search is challenging as well as significant. In this paper we propose a novel framework to exploit the users' social activities for personalized image search, such as annotations and the participation of interest groups. The query relevance and user preference are simultaneously integrated into the final rank list. Experiments on a large-scale Flickr dataset show that the proposed framework greatly outperforms the baseline.

REFERENCES

[1] J. Cui, F. Wen, and X. Tang, "Real time google and live image search re-ranking," in Proc. ACM Multimedia, 2008.

[2] ——, "Intentsearch: Interactive on-line image search re-ranking," in Proc. ACM Multimedia, 2008.

[3] N. Ben-Haim, B. Babenko, and S. Belongie, "Improving web-based image search via content based clustering," in Proc. Int'l Workshop on Semantic Learning Applications in Multimedia, 2006.

[4] W. H. Hsu, L. S. Kennedy, and S.-F. Chang, "Video search reranking via information bottleneck principle," in Proc. ACM Multimedia, 2006.

[5] A. Torralba, K. Murphy, W. Freeman, and M. Rubin, "Context- based vision system for place and object recognition," in Proc. Int'l Conf. Computer Vision, 2003.

[6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, 2005.

[7] D. Lowe, "Distinctive image features from scale-invariant key- points," International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.

[8] J. Sivic and A. Zisserman, "Video google: a text retrieval approach to object matching in videos," in Proc. Int'l Conf. Computer Vision, 2003.

[9] Y. Zhang, Z. Jia, and T. Chen, "Image retrieval with geometry- preserving visual phrases," in Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, 2011.

[10] G. Chechik, V. Sharma, U. Shalit, and S. Bengio, "Large scale online learning of image similarity through ranking," Journal of Machine Learning Research, vol. 11, pp. 1109–1135, 2010.

[11] J. Deng, A. C. Berg, and L. Fei-Fei, "Hierarchical semantic in- dexing for large scale image retrieval," in Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, 2011.

[12] K. Tieu and P. Viola, "Boosting image retrieval," International Journal of Computer Vision, vol. 56, no. 1, pp. 17–36, 2004.

[13] S. Tong and E. Chang, "Support vector machine active learning for image retrieval," in Proc. ACM Multimedia, 2001.

[14] Y. Chen, X. Zhou, and T. Huang, "One-class SVM for learning in image retrieval," in Proc. IEEE Int'l Conf. Image Processing, 2001.

[15] Y. Lu, H. Zhang, L. Wenyin, and C. Hu, "Joint semantics and feature based image retrieval using relevance feedback," IEEE Trans. on Multimedia, vol. 5, no. 3, pp. 339–347, 2003.

[16] D. Tao and X. Tang, "Random sampling based svm for relevance feedback image retrieval," in Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, 2004.

[17] D. Tao, X. Tang, X. Li, and X. Wu, "Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 28, pp. 1088 – 1099, 2006.

[18] R. Yan, E. Hauptmann, and R. Jin, "Multimedia search with pseudo-relevance feedback," in Proc. Int'l Conf. on Image and Video Retrieval, 2003.

[19] R. Yan, A. G. Hauptmann, and R. Jin, "Negative pseudo-relevance feedback in content-based video retrieval," in Proc. ACM Multi- media, 2003.