

Attribute Based Image Search Re-Ranking

Snehal S Patil¹, Ajay Dani²

¹Master of Computer Engg, Savitribai Phule Pune University, G. H. Raisonni Collage of Engg and Technology, Wagholi, Pune

²Professor, Computer and Science Dept, Savitribai Phule Pune University, G. H .Raisonni Collage of Engg and Technology, Wagholi, Pune

Abstract: *Image search reranking is an effective approach to refine the text-based image search result. Text-based image retrieval suffers from essential difficulties that are caused mainly by the incapability of the associated text to appropriately describe the image content. In this paper, reranking methods are suggested address this problem in scalable fashion. Based on the classifiers for all the predefined attributes, each image is represented by an attribute feature consisting of the responses from these classifiers. A hypergraph can be used to model the relationship between images by integrating low-level visual features and attribute features. Hypergraph ranking is then performed to order the images. Its basic principle is that visually similar images should have similar ranking scores. It improves the performance over the text-based image search engine.*

Keywords: hypergraph, attribute-assisted

1. Introduction

The existing web image search engines, including Bing, Google and Yahoo retrieve and rank images mostly based on the textual information associated with the image in the hosting web pages, such as the title and the surrounding text. While text-based image ranking is often effective to search for relevant images, the precision of the search result is largely limited by the mismatch between the true relevance of an image and its relevance inferred from the associated textual descriptions.

To improve the precision of the text-based image search ranking, visual reranking has been proposed to refine the search result from the text-based image search engine by incorporating the information conveyed by the visual modality.

The existing visual Reranking methods can be typically categorized into three categories as the clustering based, classification based and graph based methods

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters) . Initial search results from text-based retrieval can be grouped by visual closeness. In classification visual reranking is formulated as binary classification problem aiming to identify whether each search result is relevant or not. Graph based methods have been proposed recently and received increasing attention as demonstrated to be effective. The multimedia entities in top ranks and their visual relationship can be represented as a collection of nodes and edges. The local patterns or salient features discovered using graph analysis are very powerful to improve the effectiveness of rank lists.

Semantic attributes could be shape, color, texture, material, or part of objects, such as “round,” “red,” “metal,” “wheel” and “leg” etc. In this paper attribute-assisted reranking method based on hypergraph learning is introduced. First train several classifiers for all the pre-defined attributes and each image is represented by attribute feature consisting of the responses from these classifiers. Different from the

existing methods, a hypergraph is then used to model the relationship between images by integrating low-level features and attribute features. Finally, reranked list of the images obtain with respect to relevance scores in descending order.

2. Related Work

To improve the performance of searching images visual search reranking is very good option. In this section , existing visual search reranking approaches are explained along with semantic attributes and hypergraph learning.

To improve the accuracy of the text-based image search ranking, visual reranking has been proposed to refine the search result from the text-based image search engine by incorporating the information conveyed by the visual modality.

2.1 Test Based Search

When user enter query in the search engine it get related images with respect to that query in resultant image set. The search engines present today uses different image search algorithm. Basically they are text based. That mean the resultant image set contain only the images which have name identical to that query. The image set contains all the images retrieve from image database which have the name similar to input query. All this happen using text based algorithm in which ASCII values decide the ranking of characters. In database there are many images related to our query so their ranking is important to get ideal result. To rank the text based search, algorithm uses the ASCII values. As per ranking of ASCII value image names of resultant images are ranked. The main advantage of text based searching is that, it helps to get all that images from database having the name identical to our query. But disadvantage is that, it unable to concentrate on image contain. The resultant image set contain the images which not related to our search of interest, only the image name is identical to query that why they are in resultant image set. In short, text based search cannot check relevance of images. Some algorithms are their which check image relevance but they have some drawback.

Volume 4 Issue 11, November 2015

www.ijsr.net

Licensed Under Creative Commons Attribution CC BY

2.2 Content Based Image Search

It is designed to work more with actual pieces of the image. Some types use images as samples, some take various pieces of colour info, etc. Different types are there which includes, Region-based, Object-based, Example-based, and Feedback-based.

Region-based Image Retrieval:

It is low-level content-based searching. It can interpret portions of images. This works with low-level images. This can partition image and search only one portion. But this cannot work with objects. High-detail images are impossible.

Object-based Image Retrieval:

It can be works with pieces of an image, like Region-based Image IR. It can interpret images including high-detail. High-detail images are easy to search. It Can use pre-defined shapes to get images for the query. Implementation is very intense. User-interface also does not fit typical search ideas of simplicity.

Example-based Image IR:

In this users give a sample image, or portion of an image, that the system uses as a base for the search. The system then finds images that are similar to the base image. Easy for the user until the user realizes that the picture they want looks nothing like the one they have. It can be simple input for the user.

Feedback-based Image IR:

This is slightly time consuming for the user. System shows user a sample of pictures and asks for rating from the user. Using these ratings, system re-queries and repeats until the right image is found. Any image can be found with enough feedback. It may take a long time to find the image that the user wants.

2.3 Visual Reranking

It is the re-arranging of images on the basis of visual similarities. Visual reranking has been proposed to purify the search result from the text-based image search engine by incorporating the information conveyed by the visual modality.

According to the statistical analysis model used, the existing reranking approaches can roughly be categorized into three categories including the clustering based, classification based and graph based methods.

Clustering-Based Methods

Clustering analysis is very useful to estimate the inter-entity similarity. The images in the initial results are primarily grouped automatically into several many near duplicate media documents. However, for queries that return highly diverse results or without clear visual patterns, the performance is not guaranteed.

Classification-Based Methods

In the classification based methods, visual reranking is formulated as binary classification problem aiming to identify whether each search result is relevant or not. For

instance, a classifier or a ranking model is learned with the pseudo relevance feedback. However, in many real scenarios, training examples obtained via PRF are very noisy and might not be enough for training effective classifier. To address this issue, learned a query independent text based re-ranker. The top ranked results from the text based reranking are then selected as positive training examples. Negative training examples are picked randomly from the other queries. A binary SVM classifier is then used to re-rank the results on the basis of visual features.

Graph-Based Methods

Graph based methods have been proposed recently and received increased attention as demonstrated to be effective. Visual rank framework casts the reranking problem as random walk on an similarity graph and reorders images according to the visual similarities. The final result list is generated via sorting the images based on graph nodes weights. The objective is to optimize the consistency of ranking scores over visually similar samples and minimize the inconsistency between the optimal list and the initial list. Thus, the performance is significantly dependent on the statistical properties of top ranked search results. Motivated by this observation, a semi-supervised framework to refine the text based image retrieval results via leveraging the data distribution and the partial supervision information obtained from the top ranked images is proposed.

2.4 Semantic Attributes

Attributes are expected to narrow down the semantic gap between low-level visual features and high-level semantic meanings. Furthermore, the type of the most effective features should vary across queries. For example, for queries that are related to color distribution, such as sunset, sunrise and beach, color features will be useful. For some queries like building and street, edge and texture features will be more effective. It can be understood that semantic attribute could also be viewed a description or modality of image data. Using multimodal features can guarantee that the useful features for different queries are contained. Therefore, all these superiorities drive us to exploit semantic attributes for image representation in the task of web image search reranking. Based on the classifiers for all the predefined attributes, each image is represented by an attribute feature consisting of the responses from these classifiers.

2.5 Hypergraph Learning

Visual representation and semantic description are simultaneously exploited in a unified model called hypergraph. a hyperedge in a hypergraph is able to link more than two vertices.

Different from the existing methods, a hypergraph is then used to model the relationship between images by integrating low-level features and attribute features. The selection of attribute features could be conducted simultaneously through the process of hypergraph learning such that the effects of semantic attributes could be further tapped and incorporated in the reranking framework.

Graph based methods have been proposed recently and received increasing attention as demonstrated to be effective. The multimedia entities in top ranks and their visual relationship can be represented as a collection of nodes and edges. The advantage of hypergraph can be summarized that not only does it take into account pair wise relationship between two vertices, but also higher order relationship among three or more vertices containing grouping information. Regularized logistic regression trained for each attribute within each class. As attribute features are formed by prediction of several classifiers, semantic description of each image might be inaccurate and noisy.

Compared with the previous method, a hypergraph is reconstructed to model the relationship of all the images, in which each vertex denotes an image and a hyperedge represents an attribute and a hyperedge connects to multiple vertices.

In a simple graph, samples are represented by vertices and an edge links the two related vertices. Learning tasks can be performed on a simple graph. Assuming that samples are represented by feature vectors in a feature space, an undirected graph can be constructed by using their pairwise distances, and graph-based semi-supervised learning approaches can be performed on this graph to categorize objects. It is noted that this simple graph cannot reflect higher-order information. Compared with the edge of a simple graph, a hyperedge in a hypergraph is able to link more than two vertices.

3. Attribute Based Image Search Re-Ranking

3.1. Learning Scalable Discriminative Dictionary with Sample Relatedness

This method proposes a new dictionary learning method which encodes the image visual features into binary ones, and more importantly it effectively alleviates the above limitations. Our approach is motivated by the fact that humans flexibly adapt the number and nature of the attributes they use to the relatedness and variety of the observed objects, and to the complexity of the task. For example, from the great number of possible attributes to describe a set of animals, such as furry, four-legged and can swim, humans effectively only use a limited number. The principle to select attributes is simple: the chosen attributes should provide sufficient information to reflect shared and discriminative properties. This method follows this principle and combines three main things.

First, this model discovers binary features by factorizing low-level features of training images into a dictionary of arbitrary (infinite) size – the actual visual patterns present in the data form the dictionary, which adapts to the complexity of the data. The resulting Adaptive Dictionary algorithm is practical even for large data sets.

Second, this model uses the Adaptive Dictionary algorithm in a discriminative framework that not only strives for good representations, but also biases towards learning dictionary which provides discriminative binary features. In the model, the dictionary, binary representations of training samples

and classifiers are learned jointly in a max-margin framework.

Third, to enhance the generalization ability of dictionary, this method utilizes the knowledge about sample relatedness to guide the learned binary features to capture the relational structure between samples. In particular, this method encourages closely related samples to have more similar binary features than less related ones. Hence, the dictionary generalizes by exploiting related examples while still being discriminative. Figure 1.1 shows a graphical illustration of this method. The comprehensive experiments in Section 5 suggest that the resulting learned dictionary is indeed discriminative and generalizes well.

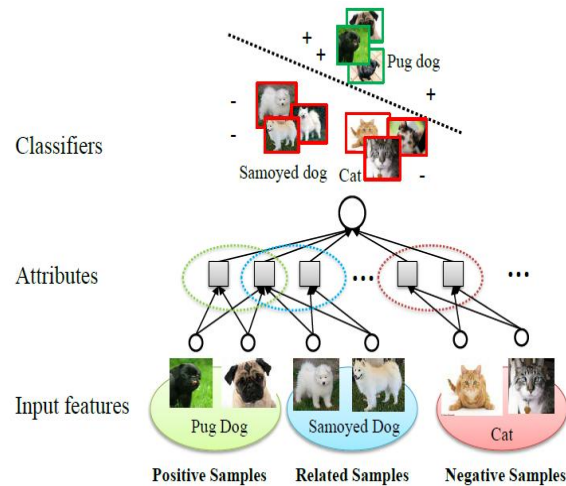


Figure 1: Illustration of the proposed dictionary learning method

It uses three types of samples for training: positive samples, samples related to the positive class and negative samples. “Attributes” of related samples (pug dog and samoyed dog) are encouraged to be shared, but the “attributes” of unrelated samples (pug dog and cat) may be different.

In short, this approach has the following benefits: (1) The size of learned dictionary automatically adapts to the complexity of the training data. Thus there is no need of bother to determine an appropriate number of basis in the dictionary as regularization parameter in this method works across a variety of data sets. (2) No need to pre-define an attribute vocabulary and tediously annotate the attributes for the training samples. (3) The model can incorporate arbitrary levels of sample relatedness from a variety of sources. In this way, the structure captured by the learned dictionary and features can be tailored to specific needs and data

3.2. Attribute-augmented Semantic Hierarchy for Image Retrieval

When a semantic hierarchy is available to structure the concepts of images, we can further boost image retrieval by exploiting the hierarchical relations between the concepts. This method presents a novel. Each semantic concept is linked to a set of related attributes. These attributes are specifications of the multiple facets of the corresponding concept. Unlike the traditional flat attribute structure, the concept-related attributes span a local and hierarchical semantic space in the context of the concept. For example,

the attribute “wing” of concept “bird” refers to appendages that are feathered; while the same attribute refers to metallic appendages in the context of “jet”. We develop a hierarchical semantic similarity function to precisely characterize the semantic similarities between images. The function is computed as a hierarchical aggregation of their similarities in the local semantic spaces of their common semantic concepts at multiple levels. In order to better capture users’ search intent, a hybrid feedback mechanism is also developed, which collects hybrid feedbacks on attributes and images. These feedbacks are then used to refine the search results based on A2SH. Compared to the attribute-based image retrieval system based on flat structure, A2SH organizes images as well as concepts and attributes from general to specific and is thus expected to achieve a more efficient and effective retrieval.

3.3 Attribute-Assisted Hypergraph Based Image Search Rearranging

Image Feature

Four types of features are useful, including color and texture, which are good for material attributes; edge, which is useful for shape attributes; and scale-invariant feature transform (SIFT) descriptor, which is useful for part attributes. A bag-of-words style feature is used for each of these four feature types. Color descriptors were densely extracted for each pixel as the 3-channel LAB values. K-means clustering represented with 128 clusters. The color descriptors of each image were then quantized into a 128-bin histogram. Texture descriptors were computed for each pixel as the 48-dimensional responses of texon filter banks. The texture descriptors of each image were then quantized into a 256-bin histogram. Edges were found using a standard canny edge detector and their orientations were quantized into 8 unsigned bins. This gives rise to a 8-bin edge histogram for each image. SIFT descriptors were densely extracted from the 8×8 neighboring block of each pixel with 4 pixel step size. The descriptors were quantized into a 1,000-dimensional bag-of-words feature. Since semantic attributes usually appear in one or more certain regions in an image, split each image into 2×3 grids and extracted the above four kinds of features from each grid respectively. Finally, obtain a 9,744-dimensional feature for each image, consisting of a $1,392 \times 6$ -dimensional feature from the grids and a 1,392-dimensional feature from the image. This feature was then used for learning attribute classifiers.

3.4 Attribute Learning

Support Vector Machine (SVM) classifier use for each attribute. However, simply learning classifiers by fitting them to all visual features often fails to generalize the semantics of the attributes correctly. For each attribute, need to select the features that are most effective in modeling this attribute. Feature selection method is apply in this case. In particular, if we want to learn a “wheel” classifier, we select features that perform well at distinguishing examples of cars with “wheels” and cars without “wheels”. By doing so, it is help the classifier avoid being confused about “metallic”, as

both types of example for this “wheel” classifier have “metallic” surfaces. Features are selected using regularized logistic regression trained for each attribute within each class, then pool examples over all classes and train using the selected features. Such regression model is utilized as the preliminary classifiers to learn sparse parameters. The features are then selected by pooling the union of indices of the sparse nonzero entries in those parameters.

For example, first select features that are good at distinguishing cars with and without “wheel”, then use the same procedure to select features that are good at separating motorbikes with and without wheels, buses with and without wheels, and trains with and without wheels. Then pool all those selected features and learn the “wheel” classifier over all classes using those selected features. In this way, effective features are selected for each attribute and the selected features are then used for learning the SVM classifier.

Attribute-Assisted Hypergraph

Attribute-assisted hypergraph learning method is used to reorder the ranked images which returned from search engine based on textual query. Different from the typical hypergraph, it presents not only whether a vertex belongs to a hyperedge, but also the prediction score that is affiliated to a specific. The weight is incorporated into graph construction as tradeoff parameters among various features. This modified hypergraph is thus able to improve reranking performance by mining visual feature as well as attribute information.

Fig. 1 illustrates the flowchart of our proposed method. After a query “baby” is submitted, an initial result is obtained via a text-based search engine. It is observed that text-based search often returns inconsistent results. Some visually similar images are scattered in the result while other irrelevant results are filled between them, such as “dog” and “disney baby”. Based on the returned images, both visual features and attribute features are extracted. In particular, the attribute feature of each image consists of the responses from the binary classifiers for all the attributes. These classifiers are learned offline. Visual representation and semantic description are simultaneously exploited in a unified model called hypergraph. Hypergraph is reconstructed to model the relationship of all the images, in which each vertex denotes an image and a hyperedge represents an attribute and a hyperedge connects to multiple vertices. The weight of each edge based on the visual and attribute similarities of images which belongs to the edge. The relevance scores of images are learned based on the hypergraph. The advantage of hypergraph can be summarized that not only does it take into account pairwise relationship between two vertices, but also higher order relationship among three or more vertices containing grouping information. Essentially, modeling relationship among more close samples will be able to preserve the stronger semantic similarity and thus facilitate ranking performance. Finally, the reranked list of the images set with respect to relevance scores in descending order.

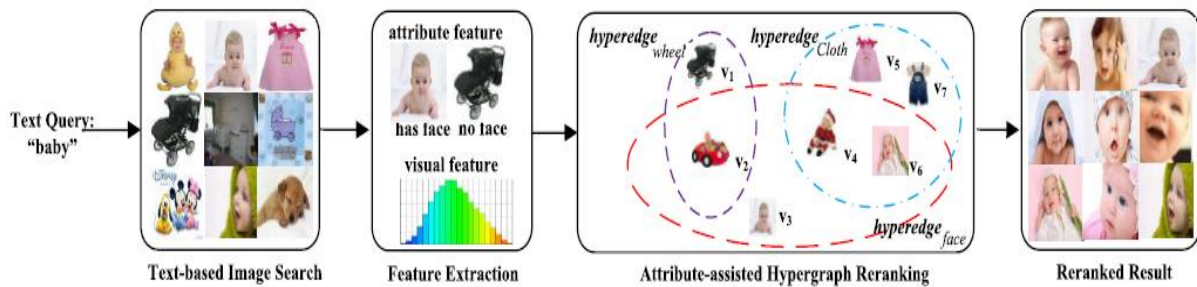


Figure 2: Flowchart of the proposed attribute-assisted hypergraph reranking method. The search engine returns the images related to the textual query “baby” and then our proposed approach is applied to reorder the result with attribute feature. The top-9 ranked images in the text-based search results and the reranked results in the first and last block are shown, respectively.

4. Conclusion

Image search reranking has been studied for several years and various approaches have been developed recently to boost the performance of text-based image search engine for general queries. This paper serves as an attempt to include the attributes in reranking framework. It is observed that semantic attributes are expected to narrow down the semantic gap between low-level visual features and high-level semantic meanings. Motivated by that, a novel attribute-assisted retrieval model for reranking images is proposed. Based on the classifiers for all the predefined attributes, each image is represented by an attribute feature consisting of the responses from these classifiers. A hypergraph can be an effective approach to model the relationship between images by integrating low-level visual features and semantic attribute features. Hypergraph ranking is performed to reorder the images, which is also constructed to model the relationship of all images.

References

[1] IEEE TRANSACTIONS ON IMAGE PROCESSING VOL. 24, NO. 1, JANUARY 2015 IEEE paper: An Attribute-Assisted Reranking Model for Web Image Search. Junjie Cai, Zheng-Jun Zha, Member, IEEE, Meng Wang, Shiliang Zhang, and Qi Tian, Senior Member, IEEE

[2] IEEE TRANSACTIONS ON MULTIMEDIA VOL.14, NO.3, JUNE 2012 IEEE paper: Prototype-Based Image Search Reranking. Linjun Yang, Member, IEEE, and Alan Hanjalic, Senior Member, IEEE.

[3] S. Changpinyo and E. Sudderth. Learning image attributes using the Indian Buffet Process. Technical report, Brown University, 2012.

[4] B. Siddiquie, R. S. Feris, and L. S. Davis, “Image ranking and retrieval based on multi-attribute queries,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2011

[5] Learning Scalable Discriminative Dictionary with Sample Relatedness. Jiashi Feng¹, Stefanie Jegelka, Shuicheng Yan, Trevor Darrell Department of ECE, National University of Singapore, Singapore Department of EECS & ICSI, UC Berkeley, USA

[6] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, “Attribute and simile classifiers for face verification,” in Proc. IEEE Int. Conf. Comput. Vis., Sep./Oct. 2009, pp. 365–372.

[7] M. Wang, L. Yang, and X.-S. Hua, “MSRA-MM: Bridging research and industrial societies for multimedia,” Tech. Rep. MSR-TR-2009-30, 2009.

[8] K. Järvelin and J. Kekäläinen, “IR evaluation methods for retrieving highly relevant documents,” in Proc. ACM SIGIR Conf. Res. Develop. Inf. Retr., 2000, pp. 41–48.

[9] W. H. Hsu, L. S. Kennedy, and S.-F. Chang, “Video search reranking via information bottleneck principle,” in Proc. ACM Conf. Multimedia, 2006, pp. 35–44.

[10] Y. Huang, Q. Liu, S. Zhang, and D. N. Metaxas, “Image retrieval via probabilistic hypergraph ranking,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2010, pp. 3376–3383.

[11] C. H. Lampert, H. Nickisch, and S. Harmeling, “Learning to detect unseen object classes by between-class attribute transfer,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2009, pp. 951–958.

[12] R. Yan, A. Hauptmann, and R. Jin, “Multimedia search with pseudo-relevance feedback,” in Proc. ACM Int. Conf. Image Video Retr., 2003, pp. 238–247.