Web Image Retrieval ReRanking with Multi-view Clustering

Mingmin Chi
School of Computer Science,
Fudan University
220 Han Dan Road, Shanghai,
China
mmchi@fudan.edu.cn

Peiwu Zhang
School of Computer Science,
Fudan University
220 Han Dan Road, Shanghai,
China
06300720175@fudan.edu.cn

Yingbin Zhao
School of Computer Science,
Fudan University
220 Han Dan Road, Shanghai,
China
zhaoyingbin@fudan.edu.cn

Rui Feng
School of Computer Science,
Fudan University
220 Han Dan Road, Shanghai,
China
fengrui@fudan.edu.cn

Xiangyang Xue
School of Computer Science,
Fudan University
220 Han Dan Road, Shanghai,
China
xyxue@fudan.edu.cn

ABSTRACT

General image retrieval is often carried out by a text-based search engine, such as Google Image Search. In this case, natural language queries are used as input to the search engine. Usually, the user queries are quite ambiguous and the returned results are not well-organized as the ranking often done by the popularity of an image. In order to address these problems, we propose to use both textual and visual contents of retrieved images to reRank web retrieved results. In particular, a machine learning technique, a multi-view clustering algorithm is proposed to reorganize the original results provided by the text-based search engine. Preliminary results validate the effectiveness of the proposed framework.

Categories and Subject Descriptors

H.3.3 [**Information Systems**]: INFORMATION STORAGE AND RETRIEVAL—*Information Search and Retrieval*

General Terms

Algorithms, Design

Keywords

Web image retrieval, reRanking, multi-view clustering

1. INTRODUCTION

Usually, there are two ways for image retrieval: (1)image-based search engine: high-dimensional, difficult for image understanding, and (2)text-based search engine: popular, easy to access and describe. For the former, content-based image retrieval (CBIR) is quite popular and successful in the last two decades [3]. However, it suffers from a so-called "semantic gap" problem between visual low-level features and semantic high-level ones.

Most of web image retrieval engines focus on the text-based index, where text queries are input to an existing web search engine, e.g., Google Image Search [1]. A huge amount of images in different classes are stored in the indexed web image database by search engine companies with labeling work. The text information usually includes the filename of a document, the block with tagging, information surrounding. However, the textual representations of images often are ambiguous and non-informative of image contents. Moreover, the query provided by user is usually short consisting of one or two terms and so the short query is more likely to be ambiguous. Therefore, returned images can include significant different semantic meanings with disorganized results.

In the poster, a web image retrieval reRanking framework is proposed to reorganize returned results with disambiguated semantic meanings. The heterogenetic contextual information is used for data analysis including both textual and visual features. In particular, multi-view clustering algorithm is proposed to reorder the initial image retrieval results provided by a text-based search engine. Preliminary results validate the effectiveness of the proposed approach.

The rest of the poster is organized as follows. A web image reRanking framework and a multi-view clustering algorithm are described in Section 2. Section 3 reports reRanking results. Finally, conclusions are given in Section 4.

2. PROPOSED FRAMEWORK

Due to disorganization and ambiguous results, it is necessary to reorganize the original retrieved images provided by the text-based search engine. In the poster, two sources of contents are integrated to reRank the original results.

2.1 Feature Extraction

In a text-based image search engine, only a single view, textual features are used for indexing. In the proposed framework, a hybrid of both textual and visual low-level features are used for data analysis.

Textual features: With image tag, webpage filename, and texts surrounding image, textual features are obtained based on a commonly used statistical measure: Term Frequency–Inverse Document Frequency(tfidf). Due to space limitations, reader is referred to [1] for details.

[1] http://images.google.com.

WWW 2009, April 20–24, 2009, Madrid, Spain.
ACM 978-1-60558-487-4/09/04.

Visual features: We used the color features. Scalable color descriptor is a color histogram in the YCrCb color space, which is encoded by a Haar transform.

2.2 Multi-View Clustering Algorithm

Two sources of contents, i.e., textual and visual features, are extracted to design the so-called multi-view clustering algorithm. In [2], each view (set) of features is separately used for clustering, and then the clustering results are combined in the end. When doing so, the clustering results by different views only have few common data points. To address this problem, we define a new similarity measure by considering both views of the features.

For textual features, the cosine similarity $\cos(\mathbf{x}_i, \mathbf{x}_j)$ is used. In order to get the same scale for visual features as that for textual ones, a normalized Euclidean distance is adopted in form,

$$\mathrm{nEuc}(\mathbf{x}_i, \mathbf{x}_j) = \parallel \mathbf{x}_i - \mathbf{x}_j \parallel / \ (\parallel \mathbf{x}_i \parallel + \parallel \mathbf{x}_j \parallel).$$

To integrate the two sources of contents, a hybrid distance measure is defined as:

$$\alpha \cdot \cos(\mathbf{x}_i^{(1)}, \mathbf{x}_j^{(1)}) + (1 - \alpha) \cdot \mathrm{nEuc}(\mathbf{x}_i^{(2)}, \mathbf{x}_j^{(2)})$$

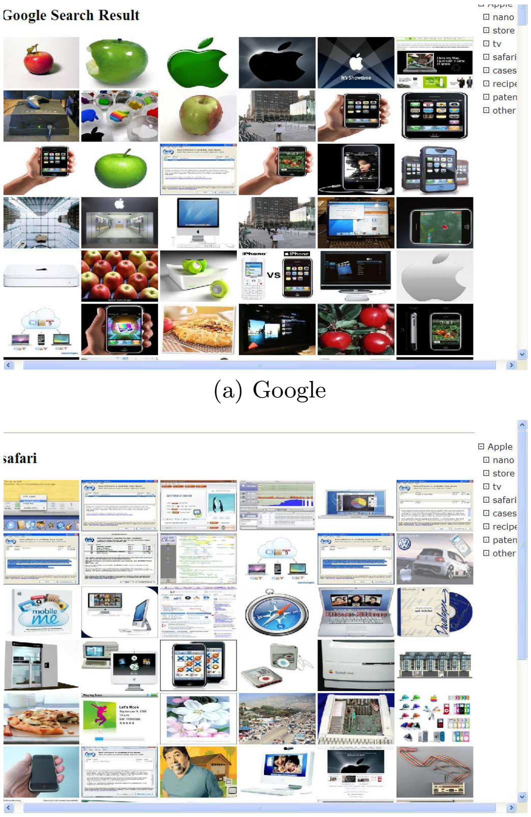
where $\mathbf{x}_i^{(1)}$ is the i^{th} pattern with textual features, $\mathbf{x}_i^{(2)}$ is the i^{th} pattern with visual features and α is a constant, which controls the contribution of textual features. If $\alpha = 0/1$, only the visual/textual features are considered and so the algorithm is reduced to a single-view clustering algorithm; otherwise, two-view features are integrated for clustering.

2.3 Web Image reRanking

After grouping the original results based on two-view features, a keyword for each cluster is obtained for reRanking according to the tfidf weight value in a given cluster. In this case, the tf weight of each word is the total frequency in the cluster instead of that in each document for text categorization. In the poster, we only use one word for the semantic meaning of an image cluster.

3. EXPERIMENTS

It is well-known that "Apple" has many semantic meanings. Thus in the experiment, the query "Apple" was used to validate the effectiveness of the proposed strategy. By Google Image Search, we obtained the results shown in Fig. 1(a) on Sept. 9, 2008. To fit the interface we defined, the left column shows the original retrieved results by Google Image Search, and the right column is a hierarchy provided by the proposed strategy, which shows the categories of different items related to "Apple", i.e., apple nano (iphone), apple store, apple tv, apple safari (browser), apple fruit (most related to recipe), apple patent, and others. From the results provided by the Google Image Search engine, if user is interested in the logo related to "patent", it is necessary for her/him to turn over too many pages to find the interesting one from the original results. However, by the proposed framework, s/he can directly click the button of "patent" shown in the hierarchical menu to find all the interesting ones. For the space limitations, only the browser (safari) is shown in Fig. 1(b) when $\alpha = 0.4$. From the results, one can see that it is much more friendly and easy for user to identify the images that s/he wants.



(a) Google

(b) Safari

Figure 1: Image reRanking results with the query "Apple" by the proposed framework, (a) the original retrieved images by Google Image Search, and (b) the reRanking images by the proposed framework.

4. CONCLUSIONS

In the poster, a web image retrieval reRanking strategy is proposed, where the images retrieved by the text-based search engine are reorganized for better visualization. In particular, a multi-view clustering algorithm is proposed to integrate two-view contents, i.e., textual and visual features extracted from web images. Experiment with the query "Apple" shows the effectiveness of the proposed framework.

5. ACKNOWLEDGMENTS

This work was supported in part by Natural Science Foundation of China under contracts 60705008 and 60873178, Shanghai Municipal R&D Foundation under contract 075107006, and Ph.D. Programs Foundation of Ministry of Education of China under contract 20070246132.

6. REFERENCES

[1] R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. Addison Wesley, 1999.

[2] S. Bickel and T. Scheffer. Multi-view clustering. In *ICDM '04: Proceedings of the Fourth IEEE International Conference on Data Mining*, pages 19–26, 2004.

[3] X. S. Zhou and T. S. Huang. Relevance feedback in image retrieval: A comprehensive review. *Multimedia Systems*, 8(6):536–544, 2003.