Telematics and Informatics 31 (2014) 477–491

Contents lists available at ScienceDirect

Telematics and Informatics

journal homepage: www.elsevier.com/locate/tele

Clustering results of image searches by annotations and visual features

Fan Wu *, Hao-Ting Pai, Yi-Feng Yan, Jeff Chuang

Department of Information Management, National Chung Cheng University, 168 University Rd., Minhsiung 62102, Taiwan

ARTICLE INFO

Article history:
Received 28 February 2013
Accepted 27 October 2013
Available online 5 November 2013

Keywords:
Iteratively structure
Category model
Text mining
Image annotation

ABSTRACT

Image on web has become one of the most important information for browsers; however, the large number of results retrieved from images search engine increases the difficulty in finding the intended images. Image search result clustering (ISRC) is a solution to this problem. Currently, the ISRC-based methods separately utilized textual and visual features to present clustering result. In this paper, we proposed a new ISRC method as called Incremental-Annotations-based image search with clustering (IAISC), which adopted annotation as textual features and category model as visual features. IAISC can provide clustering result based on the semantic meaning and visual trail; further, presented by the iteratively structure, a user can obtain the intended image easily. The experimental result shows our method has high precision that the average precision rate is 73.4%; particularly, the precision rate is 96.5% when the user drills down the intended images till the last round. Regarding efficiency, our system is one and a half times as efficient as the previous studies.

1. Introduction

Image on web has become one of the most important information for users these days. Since there are a huge number of images on the web, it is hard to find the intended images by simple query. The results of a query may contain several topics; for example, a query "Pluto" in Google Images, the results contain two different scope of semantic meaning: one is Pluto in the solar system; the other is the dog named "Pluto" in Disney world (Cai et al., 2004). It is worth noting that the results of the two scopes are mingled, that means it is hard to identify the required images immediately. Since the search engine does not clustering the results beforehand, the users have to browse all the images through page by page, and then they should manually organize these images.

Recently, web image search engine, like Google image, only returns a small fixed number of images (Wang et al., 2007). In the perspective of users, it becomes even harder to find the intended images when different scopes mingled together. To solve the problem, technique of image search result clustering (ISRC) was proposed (Cai et al., 2004; Wang et al., 2007; Ding et al., 2008; Jing et al., 2006) which is clustering images based on visual or textual features.

Regarding the visual features, such as color or texture, can be used to find similar images (Cai et al., 2004); however, the features have some limitations that it is hard to present the semantic meaning (called semantic gap) with the visual features and it takes too much time to extract the visual features from high-definitional and high-dimensional images.

Regarding the text features, it means the features of the text surrounding the image or other text sources. Text-based ISRC (Wang et al., 2007; Ding et al., 2008; Jing et al., 2006) extracted key phrases from the texts to clustering images into several clusters in a reasonable time, compared to the time to clustering images by visual features. With the text features, images in each cluster then have semantic meaning and can be used for further application. Although the images are annotated with

* Corresponding author.
E-mail addresses: kfwu@mis.ccu.edu.tw (F. Wu), greatpai@hotmail.com (H.-T. Pai).

0736-5853/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.tele.2013.10.002

semantic meanings, the semantic meaning still has its own problems, like ambiguity, homonym, insufficiency, and so on. For example, the images with "apple" annotations may contain the images about the apple computer and apple fruit; the images with "bank" annotation may contain the images about the saving bank and the river bank since the bank is a homonym.

On the other hand, the image's annotation has not enough information to fully represent the image content. In addition, the ISRC-based methods need to be performed after a query. Hence, the ISRC-based methods should process less features embedded in the images, such as complex visual features; otherwise, they cannot response immediately.

To the best of our knowledge, the state-of-the-art methods are based on ISRC (Cai et al., 2004; Wang et al., 2007; Ding et al., 2008; Jing et al., 2006); none of them integrated the textual and visual features simultaneously. Inspired by the studies (Li and Wang, 2006; Wang et al., 2008) in the area of image retrieval, adopting textual and visual features together can obtain higher accuracy because they provide images semantic meaning and visual trails. In this paper, we propose a new, called Incremental-Annotations-based Image Search with Clustering (IAISC). IAISC can extract image annotations and visual features of images in advance, and then be used to show clustering result in two ways: semantic meaning and visual pattern. Through an iterative structure, the image annotation represents the relationship among the annotations. By this way, the annotation becomes smaller and related to the images. As a result, IAISC can achieve higher precision of clustering. Although the visual feature provides good precision of annotation, it raises higher computational cost. To resolve the bottleneck, IAISC adopted the Gaussian mixture model (GMM) (Greenspan et al., 2001; Carson et al., 2002; Goldberger et al., 2006) to identify the similarity of images efficiently.

The remainder of this paper is organized in the following. In Section 2, we briefly survey the previous studies (Cai et al., 2004; Wang et al., 2007, 2008; Ding et al., 2008; Jing et al., 2006; Li and Wang, 2006; Zeng et al., 2004; Fischer and Poland, 2004; Barnard et al., 2003a,b; Chang et al., 2003; Yang et al., 2004). In Section 3, we describe the proposed IAISC. In Section 4, we make several experiments in comparison with the Google Image Swirl (GIS) (Jing and Rowley, 2012). In Section 5, we provide conclusion and future work.

2. Related work

2.1. Image search result clustering (ISRC)

ISRC (Cai et al., 2004; Wang et al., 2007; Ding et al., 2008; Jing et al., 2006) is aimed for clustering the image search result, which can use text source or visual trails as its features. By choosing the cluster, we can get the intended image in an efficient way. IGroup (Wang et al., 2007; Jing et al., 2006) is a text-based ISRC method that uses a query to search documents in page search engine, and then utilizes the PSRC (Zeng et al., 2004) algorithm to generate candidate cluster names. The high-score cluster names, applied to image search engine, are selected by pruning and merging procedure. Similar to ISRC, IGroup requires the whole documents to derive page search result. Therefore, they are inadequate for immediate request for web-based application. In addition, IGroup only takes text into consideration. In some cases, the images of the same cluster have similar semantic meaning, but they look very different.

HiCluster (Ding et al., 2008) clustering image search is constructed by two components: semantic clustering and visual clustering. Initially, it searches all documents in page search engine and uses PSRC algorithm to generate key phrases, which is used in both semantic and visual clustering. In semantic clustering, the key phrases are converted into semantic clusters by NGD calculation and K-lines clustering (Fischer and Poland, 2004). Then the semantic clusters are ranked by the semantic importance. In visual clustering, the key phrases are applied to search images in image search engine. Each image's visual feature is extracted and then ranked by the visual importance. HiCluster can provide two-layers result of clustering image search. Although HiCluster applied visual feature in ISRC, HiCluster does not succeed in fusing textual and visual clustering together. Specifically, HiCluster does not apply visual features in semantic clustering.

2.2. Page search result clustering (PSRC)

The current textual search engine, e.g., Google Images, returns a set of ranked lists for user's query. Usually, a user must go through the lists for finding out the intended website. Instead of expressing mass information, PSRC is a convenient solution by clustering the lists into varied groups. Zeng et al. (2004) proposed PSRC by changing the clustering problem into salient phrases ranking problem. Zeng et al.'s method extracts a number of valid phrases from documents, and then calculates five properties for every phrase, including: phrase frequency (the frequency of phrase shown in documents), phrase length, intra-cluster Similarity (the phrase's similarity between documents), cluster entropy (the distance of different phrases), and phrase independence (the frequency of phrase shown together with other phrase). After executing salient phrases extraction, a regression model is used to combine these properties into a single salience score. Every phrase, ranked by their score and the high-ranked phrases, represent the clusters. The object of clustering is different between in PSRC and ISRC. If an ISRC-based method plans to refer PSRC scores, it has to build complex relationship between visual features and textual features.

2.3. Auto-annotating images

By definition, image annotation is a keyword that assigned into images and converts low level visual features into high level semantic textual (Li and Wang, 2006; Wang et al., 2008; Barnard et al., 2003a,b; Chang et al., 2003; Yang et al., 2004).