Interactive Web Video Advertising with Context Analysis and Search

Bo Wang¹, Jinqiao Wang¹, Ling-Yu Duan², Qi Tian³, Hanqing Lu¹ and Wen Gao²
¹Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
²Institute of Digital Media, School of EE & CS, Peking University, Beijing 100871, China
³Department of Computer Science, the University of Texas at San Antonio, TX 78249
{bwang, jqwang, luhq}@nlpr.ia.ac.cn, {lingyu, wgao}@pku.edu.cn, qitian@cs.utsa.edu

Abstract

Online media services and electronic commerce are booming recently. Previous studies have been devoted to contextual advertising, but few work deals with interactive web advertising. In this paper, we propose to put users in the loop of collecting contextual ad information with an interaction process, establishing semantic ad links across media platforms. Given an ad video, the key frames with explicit product information are located, which allow users to click favorite key frames for searching ads interactively. A three-stage contextual search is applied to find relevant products or services from web pages, i.e., searching visually similar product images on shopping websites, ranking product tags by text aggregation, and re-search textual items consisting of semantic meaningful tags to make a recommendation. In addition, users can choose automatically suggested keywords to reflect their intentions. Subjective evaluation has demonstrated the effectiveness of the proposed approach to interactive video advertising over the Web.

1. Introduction

With the development of information technology and storage ability, many kinds of web resources are flourishing. Online ads, especially ad videos alongside the web pages, are increasing rapidly. In addition, the prosperity of electronic commerce makes online shop more popular. While more and more people enjoy the convenience and efficiency of shopping on the internet, ad videos become some kind of intermediary device to provide advertising information about products and service. In particular, when an ad video could arouse the viewers’ interest to find out or buy some relevant products, some kinds of automatic products/services recommendation would lead to effective advertising. For those ad viewers attracted by the dramatic video content, they usually hope that more details on the advertised products/services are available in a friendly way, which may lead to direct purchasing. So we propose an approach to assist users in making correct decision by providing more information on ad links between ad videos and related products ads promoted on a website. Consumers, promoter or online sellers would benefit from this recommendation.

Ad-networks like YouTube.com and Youku.com have provided an overlaying ad format for video ads, allowing advertising information being placed at the bottom fifth of the screen for 15 seconds. For the fixed ad videos setting, the overlap regions may distort close captions, logos or other visual content. This will lead to intrusive viewing experience. The main problem with in-image/video advertising is on how to reduce the annoying effects for viewers. For instance, Mei et al. [1] tried to compute discontinuity, attractiveness and visual-aural relevance to seek less intrusive points for inserting ads within video streams. Wang et al. [2] proposed to associate ad videos with relevant display ads from the Internet by matching characteristic images consisting of salient visual concepts or special logos. Advertising studies have stressed the importance of establishing ads’ relevancy for consumers [3] as irrelevant ads are usually ignored by users and relevant ads are more likely to be clicked [4]. Based on the results of image annotation, Wang et al. [5] proposed to monetize user-generated photos by modeling a user’s interest from the personal photo archive and to make targeted image/video advertising by the distribution similarity on textual topics. Wang et al. [6] [7] applied image/video content analysis to automatically link video ads with relevant Internet information, whereas the ability of visual search is insufficient to make accurate recommendation of products/services.

In this paper, we propose an interactive video advertising scheme to link ad videos with pervasive products/services information on the Internet. Interest profiles are established by capturing users’ behaviors such as clicking the interesting images or choosing keywords when they watch an ad video occasionally.
To establish the links, a three-stage contextual search is performed to search relevant products/services, which is superior to the fully automatic ad recommendation in terms of the accuracy of product/service categories.

2. Proposed Approach

Fig. 1. Overall framework.

Fig. 1 illustrates the framework of our interactive service recommendation system containing four online modules (in the colorful rectangles and lozenges): ad video analysis, user interaction, contextual search and service recommendation.

To make full use of network resources, we built a large dataset of product images from popular shopping websites, such as eBay and Amazon. In addition to product images, we stored useful tags including category, brand, price, etc. We then extracted local and global features for all stored images and created an index of local features using Spectral Hashing [8]. Video analysis is to capture the summarization images about the product. Users can select favorite one to start the contextual search. We search visually similar product images via SH based matching to coarsely find visual similar products, then the contextual text information is clustered by tag aggregation, and the results are further refined with textual re-search. As we know, content based video comprehension is not very confident and different people will like different things even though they watch the same ad video, so we would like to utilize the interactive process to give more credible and personalized recommendation. In Section 3 and 4, we will introduce these modules in detail. Section 5 will show our experiment and we conclude our work in Section 6.

3. Video Analysis

It is very critical to represent ad videos concisely for the whole searching process. The frame-by-frame matching is very time-consuming, so we select a small number of salient and representative images to summarize the commercial video. We detect the products/service related images, i.e., FMPI (Frame Marked with Production Information) images [9] to represent an ad video. As illustrated in Fig. 2, FMPI image can be regarded as a kind of document image involving graphics (e.g. corporate symbols and logos), images (e.g. products) and texts (e.g. brand names and contact information), which are usually used to highlight the advertised products, service or ideas.

The FMPI image recognizer is trained by SVM classifier. We use the probability output of the classifier to determine FMPI images and the ones with the highest probabilities within each shot of ad video are selected. For these images, we would like users choose a favorite one, such as the product appearance.

Fig. 2. Examples of FMPI images.

4. Interactive Contextual Search

As known to all, images returned from traditional web search engines are error-prone, so we try to mine matched product information with the input ad video by an interactive progressive search: visual search, tag aggregation and textual re-search.

4.1. Interactive Visual Search

Ad videos in live web TV have no tags, so the first searching step is visual search, which aims to find some visually similar images using the results of video analysis. However, when users watch an ad video, they may care different services, brands or categories of products. Although the cross-media contextual search can achieve automatic advertising [7], the results are far from satisfactory for users. For example, the system may automatically recognize that an ad video is about car, but it is very difficult to determine the car’s brand. Therefore, we need users' participation. In many cases, users are likely to get their personalized services by simply selecting their interesting key-frame.

When a representative image is chosen, several visual features are extracted for it as we did for dataset images. Global features are: Color histogram and Grid Gabor texture; local features are: SURF [10], Shape-Context [11] and Geometric-Blur [12]. For local features, we use a Naïve Bayes Nearest-Neighbor classifier [13] to search visually similar images. It is non-parametric and requires no training time.

Given a representative ad image \( Q \), our goal is to find the similar product image set \( C \):
where $p_1,\ldots,p_n$ denotes the set of stable local features in $Q$ and $f(p_i,p_j)$ measures the similarity between feature $p_i$ in $Q$ and $p_j$ in a dataset product image $I_k$,

$$f(p_i,p_j) = \begin{cases} 1, & p_j \in NN_{p_i} \\ 0, & \text{else} \end{cases}$$

where $NN_{p_j}$ means the nearest $N$ neighbors of $p_j$.

We select Spectral Hashing (SH) to accelerate the process of finding nearest neighbors. SH is a promising approach to seek compact binary codes of data-points for similarity search and it would alleviate computational cost in high-dimensional spaces to a very great extent. The basic idea is to formalize the data as a particular form of graph partitioning. By utilizing recent results on convergence of graph Laplacian eigenvectors to the Laplace-Beltrami eigenfunctions of manifolds, SH uses a spectral method to code a query point by setting threshold to a subset of eigenvectors of the graph Laplacian of similarity graph. Therefore, one can quickly determine near neighbors by hashing the query point and retrieving the elements projected into the same bins. For the fusion of different features, we adopt an entropy based fusion scheme [7].

4.2. Interactive Textual Search

Due to the gap between visual features and the abundant semantics, there will be different kinds of product images in the visual search result. To provide more specific and useful recommendation, we further make use of the context information around the product images and continue to make an interactive textual search.

We cluster the tags of visual search results into some semantically consistent classes, using the $K$-lines-based clustering algorithm [14]. There will be more results which belong to the same class as the input image. We remain top-$N$ classes of our clustering results and select keywords from these tags in the $tf-idf$ rule. Up to this step, we can regard the search process as a query expansion, for we obtain some keywords to interpret the ad video. These keywords are further used as candidates for users to select.

It is still not enough now to give accurate and personal recommendation and we need users’ participation and feedback to give information that interests them. So we provided a panel in our system to help users choose recommended keywords and input their own keywords. These two parts of keywords will be used to do the final textual search. We compute the similarity between the final keywords and each tag in the product database. The tags which have highest similarity with query keywords are regarded as results of textual re-search and they will be the automatically recommended products.

4.3. Interactive User Interface

Fig.3 shows the interactive interface of our system containing 5 modules: ad video play control, representative images selector, keywords selector, service recommendation and ad videos selector. In the example, the video is about a Blackberry cell phone. We provide 4 frames and hope users select a representative one, such as the first image. Then we show users several keywords automatically generated. Users can select what they want or input more words, such as "Curve 8520 white", to get more detailed product information they want. On the right, we provide the final products information: image, name, price and original link. Users can also get other results by clicking image links on the bottom panel. Fig.4 shows us two screenshots of the real system.

5. Experiments

The experimental data involves two parts: ad video and product database containing images and tags. We have more than 300 ad videos of 13 popular classes (car, camera, etc). Our database includes about 60000 product images and corresponding tags (including name, price, descriptions, etc.) which can be divided
into 18 categories. To evaluate the performance of interactive advertising, we conduct a subjective evaluation for our system. Five evaluators were invited to participate and each individual was allowed to randomly select ten ad videos from the database. When viewing each of the online service recommendation result, they were asked to give a score from 1 (worst) to 5 (best) based on the following aspects:

- **Relevancy.** How about the relevancy between the recommender products and video ads?
- **Acceptability.** How can you accept the recommendation?
- **Accuracy.** Does the system give right categorization for the recommender products?
- **Nonintrusiveness.** Is the recommendation intrusive for you? Do you like the recommendation if you are watching the video?

![Figure 5. Average subjective evaluation results.](image)

The average results about 30 videos are listed in Fig.5. Compared with automatic recommendation in [7], interactive advertising will improve the performance of the ad system considerably. Due to lack of textual tag for ad videos, automatic recommendation only from video content information will have some noise, even though we will conduct a three-step contextual search. Moreover, in despite of the interruption when watching videos, users would still like to get some relevant information about the commercials, especially the personalized and accurate products. The feedback also revealed that we can provide more convenient interface, such as the brand recommendation, product sorting by price, etc.

6. Conclusion

We have presented an interactive scheme for web video advertising by linking online ad videos with pervasive products/services information across the Internet. Various techniques such as video analysis, image retrieval and multimodal fusion have been applied. Subjective study has demonstrated that our approach may make more accurate ad recommendation in advertising across media. Next we will improve the performance of web video advertising by optimizing the interaction process and interface. This approach will also be extended to other videos (such as movies) with the help of semantic concept detection and the learning of user behaviors.

Acknowledgement

This work is supported by National Natural Science Foundation of China (Grant No. 60605008, 60833006 and 90920303). Also this work is partially supported by the National Basic Research Program of China under contract No. 2009CB320902.

References