

Interactive Service Recommendation Based on Ad Concept Hierarchy

Bo Wang^{1,2}, Jinqiao Wang¹, Ling-Yu Duan³, Qi Tian⁴ and Hanqing Lu¹

¹Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

²China-Singapore Institute of Digital Media, 119613, Singapore

³Institute of Digital Media, School of EE & CS, Peking University, Beijing 100871, China

⁴Department of Computer Science, the University of Texas at San Antonio, TX 78249, USA

{bwang, jqwang, luhq}@nlpr.ia.ac.cn, lingyu@pku.edu.cn, qitian@cs.utsa.edu

ABSTRACT

The rapid popularization of various online media services have attracted large amounts of consumers and shown us a large potential market of video advertising. In this paper, we aim to produce interactive service recommendation based on ad concept hierarchy by linking web videos, especially ad videos, with informative product details over the commercial websites. By introducing the domain based concept hierarchy, the recommendation quality is greatly improved. Given an ad video, we will try to semantically analyze it and conduct a contextual search from two aspects: video content and tags. For video content, we firstly extract its key frames and then make a visual search to find some relevant products. For video tags (if any) and relevant product tags gained by visual search, we will launch a textual search based on our ad concept hierarchy to judge the product category, generate some suggestion keywords, and give some recommended products to users. Users can also interactively select and adjust product categories and keywords to personalize their intentions by textual re-search. Our experimental results show that the system can successfully provide suggestion that meets the relevancy and individual requirements.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval

General Terms

Algorithms, Experimentation

Keywords

Video Analysis, Service Recommendation, Concept Hierarchy

1. INTRODUCTION

With the explosive growth in the amount of web videos accumulated from a variety of applications, everyone sees a promising market on video advertising. Ad video, as a popular and effective method to promote products and services, is

cropping up everywhere on the Internet, such as Web TV and video sharing websites. In addition, as an integral part of modern life, online shopping has grown dramatically in popularity over recent years. If video viewers are interested in the product in an advertisement, they would like to know more details or order it online immediately. Today, the average attention span of a web user is measured in seconds, so it will be very useful for the decision-making of online consumers to give them proper service recommendation. In this paper, we want to make online service recommendation with contextual search based on our product concept hierarchy in a cross-media manner.

YouTube has launched *InVideo Ads*, which is popped up in-video for 15 seconds and only take up the bottom 20% of the player screen. Despite its efforts to eliminate distraction, the problem of dampening the user experience is serious. Mei *et al.* [1] attempted to seamlessly insert video ads by computing discontinuity, attractiveness and visual-aural relevance. Wang *et al.* [2] proposed a system to automatically link video ads with relevant product information and modified the system with simple user interaction to give more accurate recommendation.

In-image advertising has the same problem as video due to the semantic gap between visual features and content interpretation by human. Wang *et al.* [3] tried to suggest relevant product information to advertise on user-generated photos by mining users' interest from personal photos. They visually analyze images by search-based image annotation and compute textual topic distributions similarity on a concept hierarchy. This hierarchy is basically a set of concepts arranged in a tree structure, which is originally used for database categorization. It is also used for many areas, such as keyword suggestion [4], due to its descriptive power on concept relations.

We propose a scheme to advertise web videos by an inter-media contextual search based on our product concept hierarchy which is considered as a topic space. For ad videos we will firstly extract their characteristic images as in [5] and match them with product photos in database. For the product tags belonging to the visually similar photos, we map them onto the topic space to get relevant product categories and generate suggesting keywords. If there are user tags for ad videos, we do the same process for them and combine these two groups of keywords. Next, we will conduct textual search using these keywords to find final automatic service recommendation results for users. Users will be very welcomed and encouraged to interactively adjust and input their own keywords.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ICIMCS'10, December 30-31, 2010, Harbin, Heilongjiang, China.
Copyright 2010 ACM 978-1-60558-840-7/09/11 ...\$10.00..

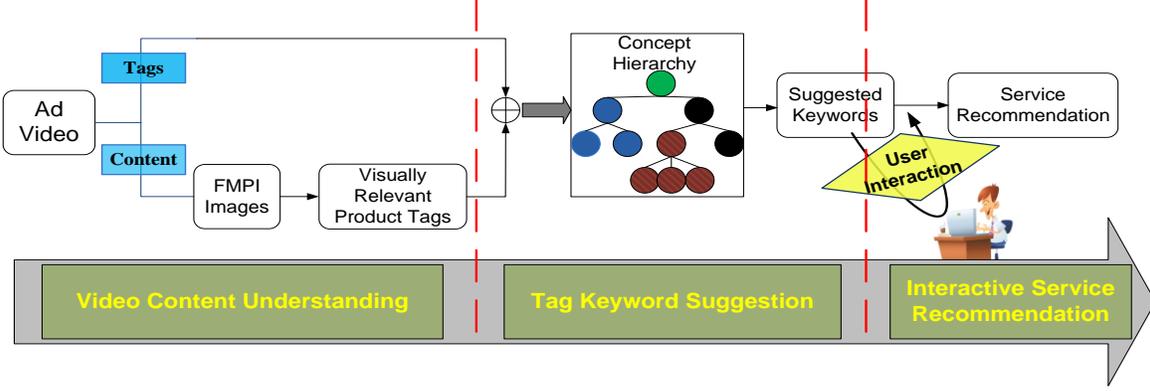


Figure 1. System framework.

2. SYSTEM FRAMEWORK

In fact, we did some preparation work for the online system. We first built a large dataset of products from popular shopping websites (eBay and Amazon) containing product images and tags (such as titles, brands, prices and shop places). Then we extracted visual features for database images and created an feature index using Spectral Hashing (SH) algorithm [6]. We learned our product concept hierarchy based on eBay website to help give relevant recommendation sequentially.

Figure 1 shows our system framework of online parts. A whole process of personalized service recommendation from video ads to product information can be divided into three steps: *video content understanding*, *tag keyword suggestion* and *interactive service recommendation*. In the first step, we capture the summarization image about the product in an ad video and search visually similar product images from the SH index of database. Then, we handle the product tags (from the result of the first step) and video tags (if any) to select some concepts or keywords, which are semantically relevant to the ad video. We aggregate these tags and map them to our concept space built offline. Users will gain some automatic recommendation and can also adjust or input their own keywords to find what they concern.

The rest of this paper is organized as follows. We will firstly introduce how to learn our concept topic space in section 3. In section 4, we will show the details for video analysis and visual image search for video content understanding. Next there are some descriptions to advertising keyword suggestion and interactive service recommendation with user interface design in section 5 and 6 respectively. At last, we show our experiment results in section 7 and conclude our work in section 8.

3. PRODUCT TOPIC SPACE

There has been some work using a hierarchical ontology to learn user interest and make advertising suggestion. The concept hierarchy involves defining concepts (meanings of concepts) and their relationships (the taxonomy structure). Many researchers [3] [4] [7] adopted the publicly available ontology provided by Open Directory Project (ODP, [8]). ODP is the world's largest manually defined web page directory, which contains thousands of categories that are hierarchically organized. Each category node of ODP can be treated as a concept, and the relationship between the categories indicated the concept relationships. ODP has extensive coverage for many topics and there are many artificially chosen web pages to be used to learn a topic or categorize a document.

But there are some problems for the abundant concepts and pages: such as noise and matching time. Many terms in the web pages are not related to the semantic concept of the target category. There are nearly a million nodes in the ODP tree, so it would cost too much time for online advertising.

Our target is for special service recommendation and coverage is limited or domain specific, therefore, the domain based concept hierarchy could be more proper for domain application. Many shopping websites has very broad products and they are carefully maintained by on-line shop owners. These sites have comprehensive and clear goods catalogue and their product tags (especially titles) are very concise for shoppers. So we derive our product concept hierarchy based on the catalog and product items from the vertical commercial websites such as Amazon and eBay.

In the ad domain, the concepts are related to each other and organized hierarchically according to their inclusion relations. A concept is described not only by itself, but also by its sub-concepts. For a concept c , whose immediate subcategories are $d_i \in S_c$, has its direct meaning $M_D(c)$ and complete meaning $M_C(c)$. We will get the following equation:

$$M_C(c) = M_D(c) \cup \{M_C(d_i) | d_i \in S_c\} \quad (1)$$

Therefore, we construct the concept hierarchy in a bottom-up way. Directly extracted phrases are firstly gathered for the leaf nodes and they are accumulated and contributed to their ancestors. Then we rank the phrases to find most relevant ones for each concept.

How to associate concepts with appropriate phrases? If a phrase is commonly used within the concept and seldom used by other concepts, we can consider it as a good one. The similarity between concept c and phrase t is measured as follows:

$$Sim(c, t) = \frac{f_c(t)}{coverage(t)} \quad (2)$$

$$f_c(t) = \frac{D_{c,t}}{D_c} \quad (3)$$

$$\frac{1}{coverage(t)} = \log\left(\frac{N}{N_t - N_{c,t} + 1} + 1\right) \quad (4)$$

In Eq. (3), $D_{c,t}$ denotes the product set within the concept c that phrase t occurs and D_c denotes all the products in c . And in Eq. (4), N_t means the number of concepts that include phrase t and $N_{c,t}$ means the number of sub-concepts of c that contains t . N is the total number of concepts. As a variation of *tf-idf*, the numerator of Eq. (2) represents the phrase frequency in the document collection associated with the product concept while the denominator indicates whether the phrase is widely used in the whole hierarchy.

4. VIDEO CONTENT UNDERSTANDING

4.1 Ad Video Analysis

As mentioned above, there are not always tags for ad videos such as commercial breaks in Web TV, and content based video analysis is a challenging and critical job. Instead of time-consuming frame-by-frame matching method, we detect the product/service related images to summarize the ad videos. Wang *et al.* [5] trained an image recognizer which can extract some FMPI (Frame Marked with Product Information) images to represent the topic of ad videos. FMPI image can be regarded as a kind of document images involving graphics (such as corporate logos), images (such as product appearances) and texts (such as brand names), which highlights the advertised products, services or ideas. Figure 2 shows us some examples of FMPI image from a Gillette shaver’s commercial video. To avoid much interruption for users and in view of the database images’ characteristic, we select one image which is more likely containing product appearance as the input of visual search.

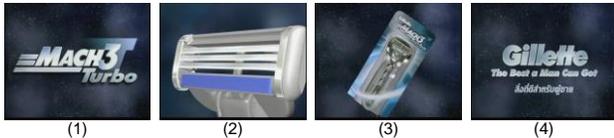


Figure 2. Examples of FMPI image

4.2 Visual Search for FMPI Image

After video analysis, we will find some visually similar product images for the related FMPI image. We firstly extracted some visual features for it as we did for database images. Some global features (Color histogram and Grid Gabor texture) and local features (SURF [9], Shape-Context [10] and Geometric-Blur [11]) are used. For local features, we apply a Naïve Bayes Nearest-Neighbor classifier to search images since it is non-parametric and requires no training time. We select the SH approach [6] to accelerate the search for nearest neighbors. SH is a promising technique to seek compact binary codes of data-points for similarity search and it can alleviate computational cost in high-dimensional spaces to a great extent. As to features fusion and selection, we adopt an entropy based fusion scheme as in [2].

5. TAG KEYWORD SUGGESTION

5.1 Tag Aggregation

The well-known semantic gap is the main problem for all content-based retrieval, so we will find different kinds of product images in the visual search results. To provide users more specific and useful recommendation, we could further make use of products’ context information, such as titles and attribute tags. For these visually relevant product tags, we cluster them into some semantically consistent categories. In general, there will be more results which belong to the same class as the input image and tags of these results are very similar to each other. Top classes of tags will be used to generate some keywords for the last textual search with the use of product concept hierarchy. We call these tags as *v-tags* (tags from ad video) compared to *u-tags* (user tags). Then we will introduce the learning process of the product concept hierarchy before keyword suggestion.

5.2 Keyword Suggestion

With the concept hierarchy in the ad domain, we perform the keyword suggestion. Firstly, we preprocess the top clusters of tags after aggregation by removing stop words and other very common

words in shopping websites, such as *buy* and *price*. Then we consider each remaining word as a query q to find the most relevant concept by mapping to the concept space. We select the most appeared concepts and their most relevant phrases as commendatory keywords.

For a certain phrase q , we use the conditional probability $P(c|q)$ to tell which concept is more important for it. According to the Bayesian theorem, we have:

$$P(c|q) = \frac{P(q|c)P(c)}{P(q)} = P(q|c) \frac{D_c}{constant} \quad (5)$$

$P(c)$ reflects the importance of the concept which can be estimated by its products number D_c . $P(c)$ can be regarded as a constant. $P(q|c)$ is measured by the formula $\log(N_{c,q} + 1)$, in which $N_{c,q}$ denotes the number of sub-concepts of c that contains t as described above.

In the case of video sharing websites, there are sometimes *u-tags* for videos which can help video understanding and hence improve the accuracy. So we will give more weights for the keywords from *u-tags* when combining with *v-tags* keywords.

6. INTERACTIVE SERVICE RECOMMENDATION

We will conduct a textual search in our database by the keywords from last step and give the preliminary service recommendation based on the search. As depicted in [2], we need user interaction to improve the system’s performance in two respects: accuracy and personality. A relevant recommendation at first sight will attract and encourage users to continue interactively use the system to get personal information. In the subsequent process, we will give dynamic keyword suggestion according to the user’s generated keywords. We calculate the similarity between a query phrase q with a phrase t as follows:

$$Sim(q, t) = \sum_{c \in C_q} Sim(c, t)P(c|q) \quad (6)$$

where C_q denotes the concepts in which the phrase q occurs. With the help of formulas in section 3 and 5, we can compute it easily.

Our system is designed in the web page format and Figure 3 shows the user interface. There are five modules: *ad video player*, *representative images*, *interactive keyword suggestion*, *service recommendation* and *ad videos selection*. In the third panel, interactive keywords suggestion, users can select some keywords generated from system or input some their own keywords to search their concerned products.

7. EXPERIMENTS

Our product database is created from eBay and Amazon and it contains more than 60,000 product images and corresponding tags (such as name, description, price and shop place) fitting into 18 types of goods. We also have 360 ad videos belonging to 13 popular product categories. The product concept hierarchy built based on the whole eBay websites includes more than 4,000 concepts. We will evaluate the performance of our system on search precision and user experience.

7.1 Experiments on Contextual Search

At first, we evaluate the system performance on final search precision using Mean Average Precision (MAP), which is the most frequently used summary measure of a ranked retrieval. Compared to the experiments in [2], we recalculate the overall MAPs for top-50 textual re-search results and find that the concept



Figure 3. User interface.

hierarchy based keyword suggestion can change slightly better from direct term frequency scheme (from 0.4 to 0.43). There are about 50 videos with user tags and we compute the effects of different textual descriptions. MAPs of three kinds of tags (“v-tags”, “u-tags” and “v-tags plus u-tags”) are: 0.42, 0.62 and 0.58. From the results we can see that, in contrast to direct term frequency method, our scheme can improve somewhat the accuracy for service recommendation; user tags can help video understanding and hence improve the precision very much.

7.2 User Study

However, accuracy is not the final goal for recommendation, and we must consider the whole performance of the system such as the friendliness and intrusiveness. So we conduct a subjective evaluation on the final interactive advertising recommendation. We invite 8 testers to participate and everyone is allowed to randomly select 20 ad videos from the database. They are asked to give a score for each recommendation result from 1 (worst) to 5 (best) based on the following aspects:

- *Relevancy*. How about the relevancy between the recommendatory products and video ads?
- *Acceptability*. How can you accept the recommendation?
- *Accuracy*. Does the system give right categorization for the commendatory products?
- *Supportiveness*. How do the suggested keywords support your interactive search?
- *Unintrusiveness*. Is the recommendation not so intrusive for you? Do you like the recommendation if you are watching the video?

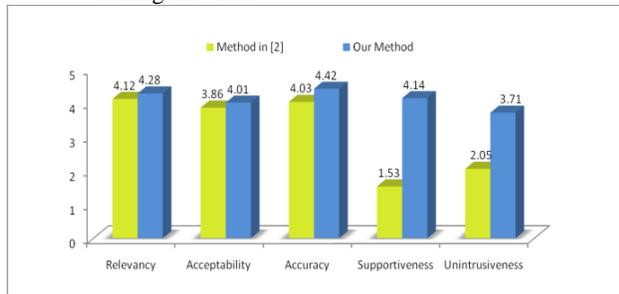


Figure 4. Comparative results of user study.

From the comparative results in Figure 4, we know that the product concept space based suggestion scheme is mostly good at supporting users in the interactive search by the relevant keywords. In addition, it has also some positive effect for recommended

results’ relevancy, accuracy and user’s acceptance. Correspondingly, it is less intrusive for users and can generate more interest for them to do more search. We also compared the interactive recommendation with automatic one, and interactive advertising will improve the performance considerably.

8. CONCLUSION

In this paper, we present a scheme for advertising recommendation based on ad concept hierarchy across the Internet. The contextual search involves video analysis, visual search, concept hierarchy based keyword suggestion and interactive textual search. Experimental results have shown that, compared with the simple term frequency method, our domain based product topic space can make some improvement for our recommendation system in several aspects, especially in supporting the interactive search and reducing distraction for users. In the future, we will extend the strategy to other kinds of videos by combining the video concepts and user tags. In addition, we will learn user interest, optimize the interaction design and modify the display manner to give more personalized recommendation.

9. Acknowledgement

This work is supported by National Natural Science Foundation of China (Grant No. 60905008, 60833006 and 90920303). Also this work is partially supported by the National Basic Research Program of China under contract No. 2009CB320902.

10. REFERENCES

- [1] Mei, T., Hua, X.S., Yang, L., and Li, S. 2007. Video Sense-Towards Effective Online Video Advertising. In Proceedings of ACM Multimedia.
- [2] Wang, B., Wang, J., Chen, S., Duan, L.-Y., and Lu, H. 2009. Semantic Linking between Ad Video and Web Service with Progressive Search. In Proceedings of ICDM Workshop on Internet Multimedia Mining.
- [3] Wang, X., Yu, M., Zhang, L., and Ma, W. 2009. Intelligent Advertising Made Possible from Users’ Photos. In Proceedings of ACM Multimedia.
- [4] Chen, Y., Xue, G.-R., and Yu, Y. 2008. Advertising Keyword Suggestion Based on Concept Hierarchy. In Proceedings of ACM WSDM.
- [5] Duan, L.-Y., Wang, J., Zheng, Y., Jin, J.S., Lu, H., and Xu, C. 2006. Segmentation, Categorization, and Identification of Commercials from TV Streams Using Multimodal Analysis. In Proceedings of ACM Multimedia.
- [6] Weiss, Y., Torralba, A., and Fergus, R. 2008. Spectral Hashing. In Proceedings of Systems.
- [7] Ma, Z., Pant, G., and Sheng, O. 2007. Interest-Based Personalized Search. In ACM Transactions on Information Systems
- [8] Open Directory Project, <http://www.dmoz.org>
- [9] Bay, H., Tuytelaars, T., and Gool, L.V. 2006. SURF: Speeded Up Robust Features. In Proceedings of ECCV.
- [10] Mori, G., Belongie, S., and Malik, J. 2005. Efficient Shape Matching Using Shape Contexts. PAMI.
- [11] Berg, A.C., and Malik, J. 2001. Geometric Blur for Template Matching. In Proceedings of CVPR.