

# Use of Content Tags in Managing Advertisements for Online Videos

Chia-Hsin Huang\*, H. T. Kung, Chia-Yung Su

Harvard School of Engineering and Applied Sciences, Cambridge, MA 02138, USA

{jashing, htk, cysu}@eecs.harvard.edu

## Abstract

*We consider the problem of managing advertisements (or ads for short) for ad-hoc online videos, such as those generated by end users. We propose a novel ad selection approach based on video tags. We first compute intrinsic values of individual video tags to each ad, called ad-specific weights of tags. The computed weight of a tag indicates the relevance of the tag to a given ad. Then based on the tag weights, we select the ad which is the most relevant to the current video which a user is viewing. Using lab experiments, we show that our tag-based approach is able to select relevant ads. For example, experiments show that 58% of the selected ads are among the top 5 most relevant ads in a pool of 64 candidate ads.*

## 1. Introduction

Recently community-generated videos, such as those end users generated video clips shared over the Web, have become increasingly popular [8]. New content of this kind is created at a fast pace and is usually not cataloged. Users generally find this unstructured content by searching through content tags provided by the community, such as content creators, service providers, and end users. To sustain such services, advertisements, or ads for short, must be considered as a potential revenue source. However, due to the ad-hoc and dynamic

---

\*Chia-Hsin Huang is with the National Taiwan University of Science and Technology. He conducted the research of this paper while visiting Harvard during 2007-08.

nature of such community-generated content, conventional static methods of selecting ads to match content are not directly applicable [4].

Traditional keyword-based ad selection approaches, which have enjoyed great success in conjunction of the operation of search engines, may not be applicable to the new environment either. This is because tags which are generated by the community do not generally convey the same high degrees of precision as keywords which are often extracted directly from a document, its metadata, or query logs. Furthermore, such keyword-based approaches could suffer from the vocabulary impedance problem [13] that advertisers and content creators use different vocabularies to describe their ads and content.

We propose in this paper a new way of using content tags to aid the selection of relevant ads. Specifically we use a regression-based method to compute intrinsic values, called *ad-specific weights*, of individual tags with respect to an ad. A tag weight indicates the relevance of the tag to the ad. Then based on the computed tag weights, a linear model is employed to select an ad which is the most relevant to the video that a user is viewing.

## 2. Tag-based Advertisement Selection

We model the tag-based ad selection problem as follows. There is a large *video set*  $\check{V} = \{V_1, V_2, \dots\}$ , which may grow indefinitely over time, and a relatively small, fixed *ad candidate set*  $\check{A} = \{A_1, A_2, \dots, A_m\}$ . We assume that each video  $V_i$  or ad  $A_j$  possesses one or more content tags in a *tag pool*  $\check{T} = \{T_1, T_2, \dots, T_n\}$ . Note that it is often the case

that advertisers associate tags with their ads.

We use a Boolean matrix  $\{\alpha_{ik}\}$  to indicate whether or not a tag is associated with a video, i.e., if  $V_i$  has  $T_k$ ,  $\alpha_{ik} = 1$ ; otherwise  $\alpha_{ik} = 0$ . Similarly, another Boolean matrix  $\{\beta_{jk}\}$  is used to indicate whether or not tag  $T_k$  is associated with ad  $A_j$ . We are interested in the ad selection problem that given any video  $V_i$  in  $\check{V}$ , select the most relevant ad from  $\check{A}$ , based on  $V_i$ 's tags.

To compute the relevance of individual tags to an given ad, we use a training video set. Without loss of generality, we assume that the *training video set* is  $\{V_1, V_2, \dots, V_x\}$  for some integer  $x$ . We further assume that we are given a *score sheet*  $\{S_{ij}\}$ , for  $i = 1, 2, \dots$  and  $j = 1, \dots, m$ . The value of  $S_{ji}$  indicates how well  $A_j$  matches  $V_i$ ; the higher the value is, the better the match. The score sheet can be generated by a survey company, a piece of software capable of estimating relevance of an ad to a video, or users' feedback.

Our *tag-based ad selection approach* consists of two phases:

**Phase 1—Weight Training Phase:** Given the training video set  $\{V_1, V_2, \dots, V_x\}$ , for each ad  $A_j$  in  $\check{A}$ , we use least squares fitting to compute ad-specific tag weights,  $W_{j1}, W_{j2}, \dots, W_{jn}$ , by minimizing the following quantity:

$$\sum_{i=1}^x \left( \sum_{k=1}^n \alpha_{ik} W_{jk} - S_{ji} \right)^2$$

If, for some  $k$ , none of the training video uses tag  $T_k$  (i.e.,  $\alpha_{ik} = 0$  for all  $i = 1, \dots, x$ ), then we set  $W_{jk} = 0$  for  $j = 1, \dots, m$ .

**Phase 2—Ad Selection Phase:** For a given video  $V_i$ , we select the most relevant ad  $A_c$  from the  $\check{A}$  as follows. For each ad  $A_j$ ,  $j = 1, \dots, m$ , we compute the following quantity  $\hat{S}_{ji}$ , called the *computed score* for  $A_j$  with respect to  $V_i$ , based on the tag weights computed in Phase 1 above:

$$\hat{S}_{ji} = \sum_{k=1}^n \alpha_{ik} \beta_{jk} W_{jk}$$

The selected  $A_c$ , for some  $c$  in  $\{1, 2, \dots, m\}$ , is the one in the  $\check{A}$  for which:

$$\hat{S}_{ci} = \max(\hat{S}_{1i}, \hat{S}_{2i}, \dots, \hat{S}_{ni})$$

We note that the computed score  $\hat{S}_{ji}$  is an approximation of  $S_{ji}$  given by the score sheet. Thus we use the maximum of these computed scores to determine the ad that we will select.

This approach can be viewed as a perceptron-like algorithm, with  $n$  (size of  $\check{T}$ ) input nodes,  $m$  (size of  $\check{A}$ ) summation nodes, and one output node that selects the maximum of sums of tag weights. The computed ad-specific weights are link weights between input and summation nodes. See [5] for a convergence proof and analysis of error bounds of perceptron algorithms.

### 3. Illustrative Example

Here we use an example to show how the tag-based ad selection approach works under the following configurations:

- The video set is  $\{V_1, V_2, \dots, V_{10}\}$ , with  $\{V_1, \dots, V_6\}$  being the training video set and  $\{V_7, \dots, V_{10}\}$  the testing video set, which will be used to evaluate the effectiveness of the ad selection approach.
- The ad candidate set is  $\{A_1, A_2, \dots, A_5\}$ .
- Table 1 shows the Boolean matrix  $\{\alpha_{ik}\}$  indicating the presence of  $i$ -th tag in video  $V_k$ . A brief description of videos and tags is parenthesized in the table.
- Table 2 is a score sheet.

The tag-based ad selection approach first calculates the ad-specific tag weights for all ads. For example, tag weights,  $\{W_{11}, W_{12}, \dots, W_{15}\}$  for  $A_1$ , are computed by minimizing the following objective function using least-squares fitting:

$$(W_{11} + W_{12} - 6)^2 + (W_{11} + W_{12} + W_{13} - 10)^2 + (W_{12} + W_{13} + W_{14} - 10)^2 + (W_{12} - 4)^2 + (W_{14} + W_{15} - 3)^2 + (W_{11} + W_{14} - 2)^2$$

We obtain the ad-specific tag weights for  $A_1$ :

$$\langle W_{11}, W_{12}, W_{13}, W_{14}, W_{15} \rangle = \langle 1.27, 4.36, 4.55, 0.91, 2.09 \rangle$$

Table 3 lists the computed ad-specific tag weights for each ad. Table 4 shows the selected ads for each

**Table 1. Video tag boolean matrix**

	$T_1$ (Car)	$T_2$ (Game)	$T_3$ (Wii)	$T_4$ (Food)	$T_5$ (Travel)
$V_1$ (Car racing)	1	1	0	0	0
$V_2$ (TV game: car racing)	1	1	1	0	0
$V_3$ (TV game: cooking)	0	1	1	1	0
$V_4$ (Boxing)	0	1	0	0	0
$V_5$ (Travel show: Food)	0	0	0	1	1
$V_6$ (Talk show)	1	0	0	1	0
$V_7$ (TV game: sports)	0	1	1	0	0
$V_8$ (Travel show)	0	0	0	0	1
$V_9$ (Show: Baby feeding)	0	1	0	1	0
$V_{10}$ (New car show)	1	0	0	0	0

**Table 2. The score sheet**

	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$
$V_1$	6	8	4	10	8
$V_2$	10	1	4	3	2
$V_3$	10	1	6	1	2
$V_4$	4	5	8	7	4
$V_5$	3	10	9	8	7
$V_6$	2	7	5	9	9
$V_7$	10	3	6	4	3
$V_8$	3	10	5	8	7
$V_9$	5	5	7	7	5
$V_{10}$	2	8	3	10	8

testing video. The tag-based ad selection approach selects  $A_1, A_2, A_3, A_4$  for  $V_7, \dots, V_{10}$ , respectively. We notice that for each testing video, the selected ad is ranked as the top 1 in the score sheet depicted in Table 2.

#### 4. Evaluation

We evaluate the performance of the tag-based ad selection approach as follows. We use a data set consisting of 64 TV ads, 300 videos, and a judge-generated score sheet with  $64 \times 300 = 19,200$  scores. The range of scores from low to high relevance is 1 through 9. For each video in the test set, we check whether or not the selected ad is ranked in the top  $X$  positions in the score sheet.

We generate the 300 videos as follows. First, we collect 87,263 videos by using “Wii” as the query term in searching YouTube. Second, to get a compact set of representative videos, we remove those videos with fewer than 120,000 views, thereby reducing the size of the video set to 488. Third, we preprocess tags by removing stopwords and per-

**Table 3. Ad-specific tag weights**

	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$
$T_1$ (Car)	1.27	3.36	0	4.82	4.36
$T_2$ (Game)	4.36	4.82	6	6.09	3.82
$T_3$ (Wii)	4.55	-7.27	-3	-8.36	-6.27
$T_4$ (Food)	0.91	3.55	4	3.73	4.55
$T_5$ (Travel)	2.09	6.45	5	4.27	2.45

**Table 4. Ads selected by the tag-based ad selection approach for the testing videos**

	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$	Selected Ad
$V_7$	8.91	-2.45	3	-2.27	-2.45	A1
$V_8$	2.09	6.45	5	4.27	2.45	A2
$V_9$	5.27	8.37	10	9.82	8.37	A3
$V_{10}$	1.27	3.36	0	4.82	4.36	A4

forming a stemming algorithm, and then we filter out infrequent tags in order to reduce the noise caused by random tag combinations. This lowers the number of distinct tags from hundreds to tens. We further remove tags that occur less than 30 times, and remove those videos that contain only 1 tag. After these steps, we obtain 422 videos. We again use view counts as a criterion to select the top 300 videos. The final video set has the following statistics: 36 distinct tags, totaling 1,243 tags. On the average, there are 4.14 tags per video with minimum 2, maximum 12, and  $\sigma^2 = 3.73$ .

We randomly select 150 out of 300 videos as the candidate training data (TD) set. The other 150 videos are used as the testing data set. Among the candidate training set, we randomly select 36, 75, or 150 videos as the actual training data (i.e., TD=36, TD=75, or TD=150) for our tag-based ad selection approach.

Figure 1 reports performance numbers. It is clear that the number of selected ads that are ranked in the top  $X$  positions will not decrease, as  $X$  increases. For instance, the tag-based ad selection approach will select an ad which is ranked as the top 1 ( $X = 1$ ) in the score sheet for about 16% of the test videos. The percentage of ads that are ranked in the top 1 through 5 ( $1 \leq X \leq 5$ ) positions increases rapidly as  $X$  approaches 5. Under TD=150,

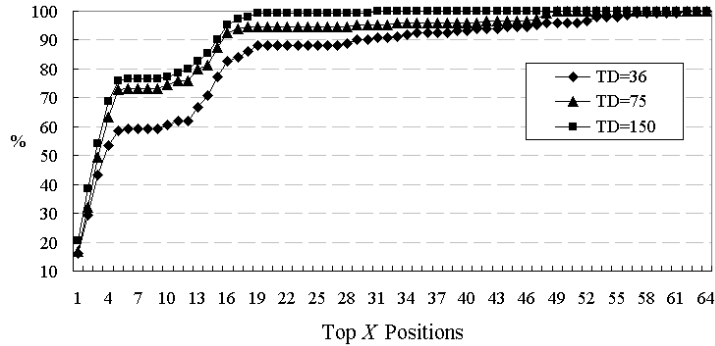


Figure 1. The rank of the ad selected by the tag-based ad selection approach

75 or 36, over 76%, 72% or 58%, respectively, of test videos are paired with a selected ad which is ranked in the top 5 in the score sheet. After  $X = 5$ , we see the increase in the percentage diminishes. This is because there are some videos that contain only few tags or contain unusual tag combinations. It is therefore much harder to select relevant ads for these videos based on their tags. Nevertheless, over 95%, 92% or 82% of videos can be paired with the selected ad that is ranked in the top one fourth positions in the score sheet while using 150, 75, or 36 training data, respectively.

We note that, as expected, the more training data is used, in general, the better performance results. The performance gain resulting from increasing the training data from 75 to 150 is not as significant as that from 36 to 75. This is an indication that 75 training videos already provides sufficient information to learn ad-specific tag weights. In fact with 36 training videos, or 12% of the total data, our tag-based method can still obtain reasonably good performance. This shows the robustness of our approach.

## 5. Discussion

We use a linear model, i.e., summing ad-specific tag weights, to derive the most relevant ad for a given video. This is similar to the use of linear models in *Artificial Neural Networks* in decision making. The linear model works better if tags are

independent, but it could still work if tags are somewhat dependent as shown in this paper. However, as claimed in [6] that tags may correspond to different information values which are not additive. This suggests future research on modeling relationships among tags.

The quality of the computed ad-specific tag weights is determined by tag combinations present in the training video set. [6, 11] claim that tagging may exhibit locality behavior. Thus, with a large training set, the linear model would capture the consensus on tagging [6, 11], which can be applied to future videos as well. Fortunately, large video sharing Web sites, like YouTube, do have a vast amount of videos available for training.

As part of this research, we have also implemented a *keyword-based ad selection approach*, which mimics popular advertising approaches employed by Web search engines. In this advertising scenario, advertisers are usually interested in maximizing the impression of their ads. Also, they may have some prior belief about which videos are suitable for pairing with their ads; however, they do not know the exact intrinsic value of video tags to their ads. Therefore, our keyword-based ad selection approach calculates *ad-specific tag frequency* of tag  $T_i$ , by summing the video frequencies of those videos which contain  $T_i$  and have scores corresponding to the ad in question greater than a certain threshold, say, 6. We compare the performance of our tag-based approach and this keyword-based

approach in terms of selecting relevant ads. We have found that our tag-based approach generally outperforms the keyword-based approach by a substantial margin. This outcome can be expected by recognizing that our tag-based ad selection approach takes ad relevance into account, while the keyword-based approaches mainly uses video frequency in selecting ads.

## 6. Related Work

There are several Internet advertising approaches. The current mainstream Web advertising approaches, like Google's *AdWords*<sup>1</sup> and Yahoo's *sponsored search*<sup>2</sup>, are more targeted because advertisers bid relevant keywords to associate with their ads. The ads are then triggered when user queries (or keywords in viewed Web pages) match the ad keywords. In social networking Web sites, users are willing to create and share personal information with others. Ads for such Web sites can be highly targeted by looking into user profiles. Myspace's *HyperTargeting*<sup>3</sup> and Facebook's *Social Ads*<sup>4</sup> are examples. Interestingly, Facebook's approach further considers social interaction between friends to achieve peer-suggested ads. Last but not the least, Baidu proposes *Brand-Link*<sup>5</sup>, a brand-based advertising approach that could generate brand-related content for brand-related search queries. However, details are still unavailable in the literature. To the best of our knowledge, none of the above approaches have been applied to community-created video Web sites. Also, none use content tags for aiding the ad selection process. One reason may be that they need significant modifications to be applicable to such an environment. It remains to be seen whether these approaches are effective in advertising for online ad-hoc videos.

Specifically, in search engines' keyword-based advertising approaches, the quality of selected ads

is determined by advertisers' willingness to bid. Several methods have been proposed for choosing relevant keywords [9] or cheaper but still effective keywords [1]. In general, this advertising approach suffers from the vocabulary impedance problem [13]. That is, an ad may not be matched to related queries or content because advertisers and content creators use different vocabularies. The work in [13] tries to solve this problem by including additional ad keywords from other related Web pages. However, increasing ad keywords may be impractical because it also increases the cost of advertising and the effort related to managing a large number of keywords. Our tag-based ad selection approach can identify tag relevance, and uses the information to associate ads. Thus, our tag-based approach, in contrast to the keyword-based approach, is less likely to suffer from the vocabulary impedance problem.

In fact, community-generated tags have been suggested to improve the quality of search results [2, 14] and to generate personalized tags for Web pages [3]. We believe that community-generated tags can also improve the efficiency of advertising. However, the use of tags to select ads, to our knowledge, has not yet been studied prior to this paper.

Several research projects have focused on modeling and analyzing the characteristics of social tagging systems. [11] uses a "user-tag-resource" model of social tagging and uses Flickr as a case study. [10, 11] find that content tags can indicate more information than the content. [7], on the other hand, examines whether the frequency of tags for popular sites with a long history can be described by a power law distribution. These findings suggest that tags can introduce information more than the content itself. These results may be viewed as support of our approach described in this paper. Furthermore, [12] claims that there are relationships among tags which cannot be represented by tag clouds. Also, [7] evaluates information value of a tag based on a cosine distance measure, which captures a degree of co-occurrence. They suggest that tag combinations may have different information values that are not additive. It seems that tags may have some correlations and thus the information

<sup>1</sup><http://adwords.google.com/>

<sup>2</sup><http://searchmarketing.yahoo.com/srch/>

<sup>3</sup><http://www.myspace.com/hypertargeting>

<sup>4</sup><http://www.facebook.com/business/?socialads>

<sup>5</sup><http://www.reuters.com/article/pressRelease/idUS124385+03-Dec-2007+PRN20071203?symbol=BIDU.O>

tags carry are most likely not equal. As mentioned early, these results suggest a future research direction on extending our linear model of tag weights.

## 7. Conclusions

We summarize important points of this paper:

- We describe a novel ad selection approach for ad-hoc online videos. The method targets a user by selecting the most relevant ad based on the tags of the video the user plays. To the best of our knowledge, this is the first time video tags have been applied to ad selection problems for online videos.
- We propose a method of computing intrinsic values, called *ad-specific weights*, of individual tags with respect to an ad. The computed weight of a tag indicates the relevance of the tag to a given ad. Thus, the computed tag weights may aid advertisers to determine relevant tags for their ads and the amount to bid for these tags. Thus, our approach, in contrast to Web search engines' keyword-based approaches, can alleviate the vocabulary impedance problem [13].
- We have designed a method that can learn the relative significance of various tags in a given video set. The method takes advantage of the existence of a very large volume of videos (e.g., those available from YouTube and other video sharing sites) that can be used in the learning process.
- Under a lab experimental setup, we have demonstrated that our tag-based ad selection approach is able to select relevant ads.

## References

- [1] V. Abhishek and K. Hosanagar. Keyword generation for search engine advertising using semantic similarity between terms. In *Proc. of the 9th international conference on Electronic commerce*, pages 89–94, 2007.
- [2] S. Bao, G. Xue, X. Wu, Y. Yu, B. Fei, and Z. Su. Optimizing web search using social annotations. In *Proc. of the 16th international conference on World Wide Web*, pages 501–510, 2007.
- [3] P. A. Chirita, S. Costache, W. Nejdl, and S. Handschuh. P-tag: large scale automatic generation of personalized annotation tags for the web. In *Proc. of the 16th international conference on World Wide Web*, pages 845–854, 2007.
- [4] X. Dreze and F. Zufryden. Internet advertising: Is anybody watching? *Journal of Advertising Research*, 17(4):60–79, 1999.
- [5] Y. Freund and R. E. Schapire. Large margin classification using the perceptron algorithm. *Machine Learning*, 37(3):277–296, 1999.
- [6] S. Golder and B. A. Huberman. The structure of collaborative tagging systems. Technical report, HP Labs, 2005.
- [7] H. Halpin, V. Robu, and H. Shepherd. The complex dynamics of collaborative tagging. In *Proc. of the 16th international conference on World Wide Web*, pages 211–220, 2007.
- [8] M. J. Halvey and M. T. Keane. Analysis of online video search and sharing. In *Proc. of the 18th conference on Hypertext and hypermedia*, pages 217–226, 2007.
- [9] A. Joshi and R. Motwani. Keyword generation for search engine advertising. In *Proc. of the 6th IEEE International Conference on Data Mining*, pages 490–498, 2006.
- [10] M. E. Kipp and D. G. Campbell. Patterns and inconsistencies in collaborative tagging systems: An examination of tagging practices. In *Proc. of the Annual General Meeting of the American Society for Information Science and Technology*, 2006.
- [11] C. Marlow, M. Naaman, D. Boyd, and M. Davis. Ht06, tagging paper, taxonomy, flickr, academic article, to read. In *Proc. of the 17th conference on Hypertext and hypermedia*, pages 31–40, 2006.
- [12] E. Michlmayr and S. Cayzer. Learning user profiles from tagging data and leveraging them for personal(ized) information access. In *Proc. of the WWW '07 Workshop: Tagging and Metadata for Social Information Organization*, 2007.
- [13] B. Ribeiro-Neto, M. Cristo, P. B. Golgher, and E. S. de Moura. Impedance coupling in content-targeted advertising. In *Proc. of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 496–503, 2005.
- [14] Y. Yanbe, A. Jatowt, S. Nakamura, and K. Tanaka. Can social bookmarking enhance search in the web? In *Proc. of the 2007 conference on Digital libraries*, pages 107–116, 2007.