A Broadcast Model for Web Image Annotation

Jia Li[1], Ting Liu[1], Weiqiang Wang[1,2], and Wen Gao[1]

[1] Institute of Computing Technology, CAS, Beijing 100080, China
[2] Graduate University of Chinese Academy of Sciences (GUCAS)
{jli, tliu, wqwang, wgao}@jdl.ac.cn

Abstract. Automatic annotation of Web image has great potential in improving the performance of web image retrieval. This paper presents a Broadcast Model (BM) for Web image annotation. In this model, pages are divided into blocks and the annotation of image is realized through the interaction of information from blocks and relevant web pages. Broadcast means each block will receive information (just like signals) from relevant web pages and modify its feature vector according to this information. Compared with most existing image annotation systems, the proposed algorithm utilizes the associated information not only from the page where images locate, but also from other related pages. Based on generated annotations, a retrieval application is implemented to evaluate the proposed annotation algorithm. The preliminary experimental result shows that this model is effective for the annotation of web image and will reduce the number of the result images and the time cost in the retrieval.

Keywords: Web image annotation; retrieval; Broadcast Model.

1 Introduction

Content-based web image retrieval [4] is quite popular now. Images on the Internet usually can be annotated through the cues from other information. Some existing algorithms [2,7] first divide a web page into blocks through a certain page segmentation algorithm, such as [1,5,6] to make each block contain a single topic. Then the image in the block is annotated through the cues from ALT (alternate text), texts and links in that block. But these algorithms ignore the content of the block while measuring its importance. Sometimes appropriate words for annotating an image can't be found within the page where the image locates.

To overcome these disadvantages, we present a Broadcast Model (BM). In this model information used to annotate an image can locate at anywhere in a website, and the importance of each block is calculated with respect to their content.

In the remaining parts, Section 2 gives a clustering method for page segmentation. Section 3 describes the proposed BM and presents an iterative implementation for BM. Section 4 gives experimental results. Section 5 concludes the paper.

2 Page Segmentation

In our algorithm, a web page is first divided into blocks through the following clustering procedure:

Step 1: Extract all the basic elements in a web page, such as images, links or texts, which can be identified by different tags <img>, <a> <#text> respectively. The size and position for each basic element is directly recorded.

Step 2: Images with a large enough area and a proper width-height-ratio are considered to be valid.

Step 3: Cluster basic elements into rectangle blocks based on the following rules.

1) Each element is considered as a block. A block B_i contained by another block B_j will be combined into it.
2) For two blocks that don't both contain valid images, they will clustered to a new block if the new rectangle block does not intersect with other existing blocks.

Step 4: for each block B_k , evaluate its importance $I(B_k)$ by formula (1):

$$I(B_k) = WA(B_k) / \sum_{j=1}^{N} WA(B_j) \ , \tag{1}$$

$$WA(B_j) = W^{valid} \times S_j^{valid} + W^{invalid} \times S_j^{Invalid} + W^{link} \times S_j^{link} + W^{text} \times S_j^{text} \ , \tag{2}$$

Where S_j^{valid} , $S_j^{invalid}$, S_j^{link} , S_j^{text} denotes the total area of valid images, invalid images, links, and texts in block B_j , and W^{valid} ,$W^{invalid}$,W^{link} ,W^{text} denote the weights of the corresponding basic elements. Generally we have $W^{valid} > W^{text} \geq W^{link} > W^{invalid}$. Fig.1 gives an example of page segmentation result.

Fig. 1. An example of page segmentation

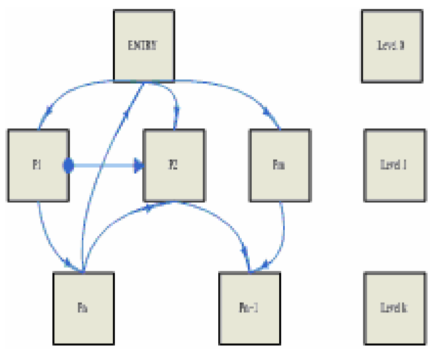

Fig. 2. A structure graph of a website

3 Broadcast Model (BM)

"Broadcast" here means an interaction between blocks and web pages. In this section, we first give an overview of the proposed broadcast model in subsection 3.1, and then present an iterative implementation of BM.

3.1 Overview

For pages in a website, they are organized together through hyperlinks. We use $S \xrightarrow{1} T$ to denote a hyperlink from an object S to another object T . Objects here can be a page or a block. Correspondingly, $B_i \xrightarrow{N} P_{jN}$ denotes a path from a block B_i to a page P_{jN} through N leaps. The structure of a website is illustrated in Fig. 2. The entry in Fig.2 denotes the first page of the website and levels are extracted from the address of pages. Generally a page P_m in low level tends to be more general in content than P_n in high level. In the model, the content in a block B_i is only related with information in the pages in higher levels. The information correlation reduces as the distance between B_i and the pages increases. Normalized Gaussian function will be used to model this correlation as in formula (3):

$$E(x) = \frac{G(x)}{G(0)} \quad G(x) = \frac{1}{\sqrt{2\pi}} \exp[-\frac{x^2}{2}] \quad x = 0,1,2,3...... \tag{3}$$

$E(N)$ shows the information correlation between a block and a page N steps away from it. When $x \geq 3, E(x) < 0.011$. So only information in pages one or two steps away from a block are used when annotating an image in the block in our model.

For an object (a page or a block), we use a feature vector $F = [<W_1, D_1>,...,<W_i, D_i>,...,<W_N, D_N>]$ to characterize its content. Here $<W_i, D_i>$ $i = 1,2,...N$ denote a keyword and its descriptive factors (DF), which show the importance of a keyword for describing the content of the object. We use $D_j^{'} = D_j / \sum_{i=1}^{N} D_i$ to normalize a feature vector and let $D_1 \geq D_2 ..... \geq D_N$. In our model, the max length of the feature vector for a page is longer than that for a block, since we usually need more words to characterize the content of a page than a block. For valid images in a block, we use keywords and their importance in the feature vector for the block to annotate them. An iterative process is developed to evaluate the values of feature vectors for different blocks or pages.

3.2 An Iterative Algorithm to Implement BM

The proposed algorithm to implement BM assumes that most information useful for annotating an image locates in nearby pages or ALT/surrounding texts/links in the block the image belongs to.

3.2.1 Initialization

After pages are divided into blocks and the importance of each block are calculated, we start up the annotation of images through the following initialization procedure.

Step 1: For each block, add keywords and their occurrences from surrounding texts, links and ALT (if any) into the feature vector of that block with different weights, since their importance in the annotation are different. The information in a link usually summarizes a topic about the page it links to. So keywords and their frequencies associated with the link are also added to the feature vector of the page it links to.

Step 2: For the feature vector of each block, the first N_b keywords w_i with largest weighted frequencies f_i ($i = 1, 2, ..., N_b$) are reserved. The descriptive factor for w_i is calculated by $D_i = f_i / \sum_{n=1}^{k} f_n$, where k is the length of the feature vector, $k \le N_b$.

Step 3: For each page, keywords from outside links and all the blocks in the page are sorted into a list according to their frequencies which are further weighted by the importance of the block it belongs to before sorting. The first N_p keywords and keywords from outside links are used to represent the content of the page. Then normalize the feature vector of each page.

3.2.2 Annotating Valid Images

In the iterative procedure, the feature vector of each block is updated to make it better summarize the topic in the block, i.e., determine the suitable keywords and their proper DF values for the description of the topic. The details of the iteration algorithm are summarized as follows:

Step 1: Set the current iteration times Time=0, and the total iteration times T.
Step 2: Process all the blocks and pages:

For each page P_i :

(a) For each block B_j in page P_i which contains valid images, let $S = [< W_1, D_1 >, ..., < W_i, D_i >, ..., < W_N, D_N >]$ denote its feature vector.

1) Calculate two thresholds $T_H = \sum_{i=1}^{N} D_i / N$ and $T_L = \beta T_H$, $\beta \in (0,1)$. Based on these two thresholds, the feature vector S is divided into three parts. Let $D_1 \ge ... \ge D_{N1} \ge T_H \ge D_{N1+1} ... \ge D_{N2} \ge T_L \ge D_{N2+1} ... \ge D_N$ and a metrics Ω is defined in formula (4) to measure how well S is for the description of the topic.

$$\Omega = (\sum_{i=1}^{N1} D_i + \frac{1}{2} \sum_{i=N1+1}^{N2} D_i) / (\sum_{i=N2+1}^{N} D_i + \frac{1}{2} \sum_{i=N1+1}^{N2} D_i) , \quad (4)$$

Then Normalize S and calculate the Ω value based on it.

2) Calculate the feature vector S_{out} representing information from other pages useful for annotation. It is obtained through weighting all the feature vectors of the pages one or two steps away from B_j with $E(0)$ and $E(1)$ respectively. Only the first N_p keywords in S_{out} with largest weighted DF value are reserved. Then S_{out} is normalized.

3) Modify feature vector S using S_{out} through the following three steps:
Enhance: For each keyword W_k whose DF value D_k is less than T_H , if W_k also appears in S_{out} with DF value $D_k^{'}$, D_k in S is replaced with $D_k + D_k^{'}$ when this change can increase the Ω value of a new S . Since the words with DF values greater than T_H are considered to be already proper for annotation, we don't enhance them in the same way.
Reduce: Normalize S and omit words whose DF values are less than T_L .
Disturb: Calculate the new Ω value based on S . If the number of keywords in S is low or the change of Ω values after the current iteration is lower than a predefined threshold, the keyword (with the smallest DF value in S_{out}), which doesn't appear in S , is added into S . That means we add disturbance into S from gentle to great in order to reach a global optimization and try not to induce the over-unsteadiness of S . Then S is normalized again.

(b) Modify the feature vector for the page P_i based on the change of feature vectors of all the blocks in it.

Step 3: If Time < T then makes Time=Time+1 and go to step 2. Otherwise end the iterative procedure and take the feature vector in each block as the annotation of images in them.

The goal of the iterative procedure is to make the value of Ω for the feature vector of each block containing valid images as high as possible, and at the same time to keep a proper number of keywords in the feature vector. After the iteration, the feature vectors of blocks will contain a proper number of words and proper DF values to denote the topic in the block. Then they will be considered to be more effective as the annotation of the image in this block.

4 Experiments

In our experiment 15890 pages are crawled from http://www.nationalgeographic.com, and the number of images is 20189. In the experiment, the area of valid images is required to exceed 6000 pixels and width-height-ratio for valid images is between 0.2 and 5. $w^{valid} = 20, w^{link} = 1, w^{invalid} = 1$ and W^{text} is set with respect to text properties, such as font, font size and background. In the iterative procedure, we set $N_B = 10, N_p = 10$, $\beta = 0.5$. A text-based image retrieval platform is used to evaluate the effectiveness of generated annotation for images. In the retrieval system, a user can query images with several words that form a query vector Q with equal DF values initially. An image will gain a score increase by $D_X \times D_Y$ if a keyword W_i in Q is one of its annotation keywords, where D_X and D_Y are their DF values in the annotation of the image and the query vector Q respectively. After that the images with positive scores are shown to the user.

200 valid images are randomly selected and described with three keywords by a user at most. These images are retrieved and the results are used to evaluate the annotation. We will evaluate our model through three aspects: rank of the expected

images in the result, average time cost in the retrieval process and average number of images generated in the retrieval. A better annotation will cause small ranks, little average time cost and small average number of results generated in the retrieval. To make a comparison we also check these three aspects in the result without using the BM. The percentages of results whose rank in the retrieval result are smaller than specific values are shown in Fig. 3.

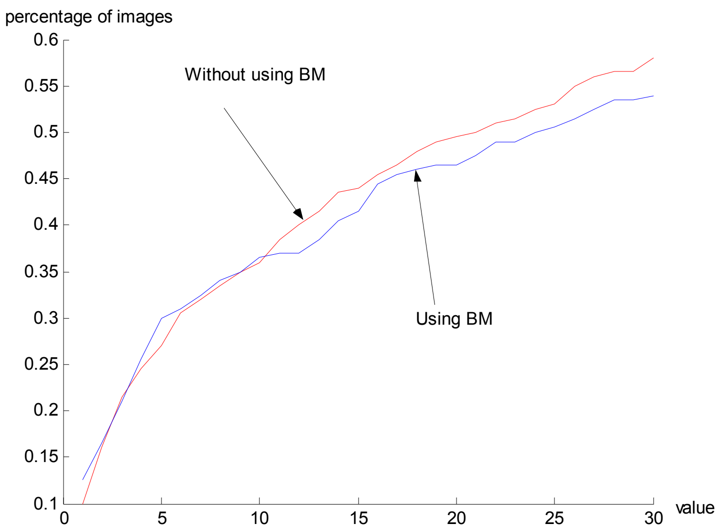

Fig. 3. Percentage of result image in user's tolerance range

Further analysis reveals that the percentage of images with ranks less than 30 are almost the same whether using BM or not. But when using BM, the average time used in the retrieval is about 10% less than the average time cost in the retrieval without using BM. And the average number of result images whose score is positive in the retrieval is also 17% less than the average result number without using BM. That is because the length of feature vectors used for the annotation of images is shorter and only eminent words are left. The experiment shows that when using BM, the time cost in the retrieval and the average number of result images have a certain reduction while ranks of the expected images change a little.

5 Conclusions

In this paper we present a method to evaluate the importance of a block after the page segmentation procedure. Based on the segmentation algorithm we propose a broadcast model algorithm to choose valid annotation for images on the Internet from associated information. In this model the source of the keywords for annotating an image is not limited to a web page where the image locates. In the future we will extend our work to a larger set of web pages and low-level features will be used for the annotation of images.

References

[1] Deng Cai, Shipeng Yu, Ji-Rong Wen, Wei-Ying Ma. VIPS: a vision-based page segmentation algorithm. Microsoft Technical Report, MSR-TR-2003-79, 2003

[2] Deng Cai, Xiaofei He, Wei-Ying Ma, Ji-Rong Wen and Hong-Jiang Zhang, Organizing WWW Images Based on the Analysis of Page Layout and Web Link Structure,2004 IEEE International Conference on Multimedia and Expo., Taipei, Jun. 2004.

[3] Huamin Feng, Rui Shi, and Tat-Seng. Chua. A bootstrapping framework for annotating and retrieving WWW images. ACM International Conference Multimedia 2004, pages 960-967, 2004.

[4] M. L. Kherfi, D. Ziou and A.Bernadi, Image Retrieval from the World Wide Web: Issues, Techniques and Systems. ACM Computing Surveys, vol 36, no. 1, March 2004.

[5] S. Chakrabarti. Integrating the document object model with hyperlinks for enhanced topic distillation and information extraction. In Proc. of the 10th International Conference on World Wide Web, pages 211--220, 2001.

[6] Shian-Hua Lin, Jan-Ming Ho. Discovering Informative Content Blocks from Web Documents. In Proc. of the SIGKDD'02 Conf., pages 588--593, 2002

[7] S. P .Li, S. Huang, G. R.Xue, Y. Yu," Block-based language modeling approach towards web search " In Proc. of the seventh Asia-Pacific Web Conference (APWeb2005), March 29 - April 1, 2005, Shanghai, China.