PageSense: Toward Stylewise Contextual Advertising via Visual Analysis of Web Pages

Tao Mei, *Senior Member, IEEE*, Lusong Li, Xinmei Tian, *Member, IEEE*, Dacheng Tao, *Fellow, IEEE*, and Chong-Wah Ngo

***Abstract*— The Internet has emerged as the most effective and a highly popular medium for advertising. Current contextual advertising platforms need publishers to manually change the original structure of their Web pages and predefine the position and style of embedded ads. Although publishers spend significant effort optimizing their Web page layout, a large number of Web pages contain noticeable blank regions. We present an innovative stylewise advertising platform for contextual advertising, called PageSense. The "style" of Web pages refers to the visual appearance of a Web page, such as color and layout. PageSense aims to associate style-consistent ads with Web pages. It provides two advertising options: 1) If publishers predefine ad positions within Web pages, PageSense will analyze the page style and select ads, which are consistent with the Web page layout, and 2) if publishers impose no constraints for ad placement, PageSense will automatically detect blank regions, select the most nonintrusive region for ad insertion, associate color-consistent ads with the Web pages, and deliver them to blank regions without breaking the original Web page style. Our experiments have verified the effectiveness of PageSense as a complement to existing contextual advertising.**

***Index Terms*— Contextual advertising, multimedia advertising, online advertising, visual content analysis.**

I. Introduction

OVER the decades, the advertising industry has grown consistently and rapidly. In an advertising system, there is usually an intermediary commercial ad-network (i.e., a service provider between the publisher and the advertiser) in charge of optimizing the ad selection and display, with the twin goals of increasing revenue (shared by the publisher and ad-network) and improving user experience. With those goals, it is preferable to the publishers and profitable to the advertisers to have ads relevant to media content rather than generic ads [1]. By implementing a solid advertising strategy into an existing content delivery chain, both the publisher and the advertiser have the ability to deliver compelling content, reach a growing online audience, and eventually generate additional revenue from online media.

Across the Web, there are many forms of Web advertising, such as pop-up ads, banner ads, and floating ads. However, these ads are presented in an aggressive manner, and may partially cover the essential content of a Web page, leading to an adverse effect as most users will immediately close all of them. Furthermore, these ads are not likely to attract user attention, since they are irrelevant to the Web pages. Compared with the above online ads using the costly machine-gun strategy and presenting in a direct and often in an aggressive way, contextual advertising dynamically displays text and graphic ads relevant to Web pages [2]. It is believed that these advertisements have a better chance of attracting users, because they share a similar context to that of user search and browse intention [3]. Therefore, contextual advertising has become an effective method for online advertising. The most popular contextual advertising network is Google's AdSense. It displays text, image ads, or link units on websites that are targeted to the Web page content.[1] Despite these advantages, contextual advertising usually requires manual work and suffers from the following limitations.

1) Existing contextual advertising systems require publishers to rearrange the layout of their Web pages and reserve a space for placing ads, as shown in Fig. 1. Just like other elements of the Web page, e.g., images, publishers have to manually set the position of ads by using HTML tags. This is not an easy task for average users who are not familiar with HTML to modify their Web pages to embed ads in proper positions. Moreover, the publishers usually do not want to spend too much effort optimizing their Web page layout for ad placement. It would be better if ad-networks could automatically find suitable ad positions to save publishers' efforts.

Manuscript received November 24, 2015; revised March 31, 2016, May 25, 2016, and July 7, 2016; accepted July 28, 2016. Date of publication August 10, 2016; date of current version January 5, 2018. This work was supported in part by the National Science Foundation of China under Contract 61572451 and Contract 61390514, in part by the Youth Innovation Promotion Association CAS under Grant CX2100060016, in part by the Fok Ying Tung Education Foundation, in part by the Fundamental Research Funds for the Central Universities under Grant WK2100060011, and in part by the Australian Research Council Projects under Grant FT-130101457 and Grant LP-140100569. This paper was recommended by Associate Editor R. Hamzaoui. *(Corresponding author: Xinmei Tian.)*

T. Mei is with Microsoft Research Asia, Beijing 100080, China (e-mail: tmei@microsoft.com).

L. Li is with Graymatics Inc., Singapore 118223 (e-mail: lilusong@gmail.com).

X. Tian is with the CAS Key Laboratory of Technology in Geo-spatial Information Processing and Application System, University of Science and Technology of China, Hefei, 230027, China (e-mail: xinmei@ustc.edu.cn).

D. Tao is with the Centre for Quantum Computation and Intelligent Systems, Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW 2007, Australia (e-mail: dacheng.tao@uts.edu.au).

C.-W. Ngo is with the Department of Computer Science, City University of Hong Kong, Hong Kong 999077 (e-mail: cwngo@cs.cityu.edu.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2016.2598702

[1] www.google.com/adsense

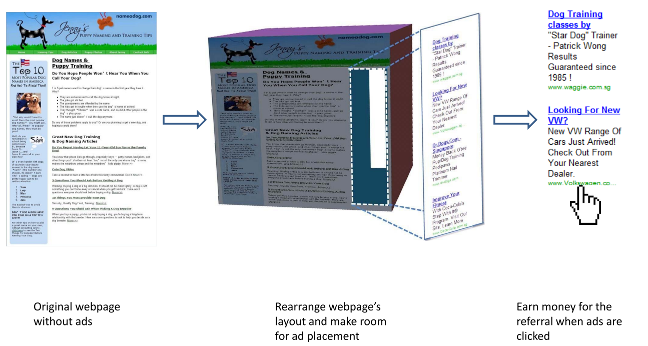


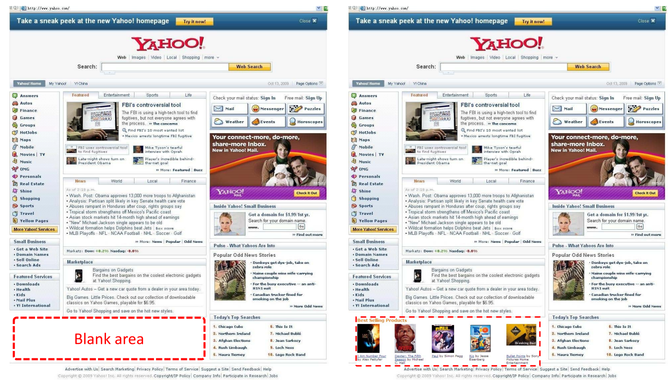
Fig. 1. Website publisher rearranges the layout of a Web page to make room for placing AdSense ads. If the visitor clicks, one of the AdSense ads served to the Website, and the Website owner is credited for the referral.

(a) (b)

Fig. 2. AdSense examples. (a) Good example. (b) Bad example.

2) Intuitively, ads relevant to the page style will reduce visual intrusiveness. Current contextual advertising allows publishers to customize the appearance of ads (e.g., links and text descriptions) to match the hosting Web page's appearance. Fig. 2(a) shows a good example of customizing AdSense ads. The links on the site are all red and the text is white, while the Website publisher uses the same red links and white text in their ads as well. However, this significantly increases publishers' effort on manually deciding the placement and style of ads. Moreover, current contextual advertising does not consider whether the visual contents match the style of the hosting page, as shown in Fig. 2(b). We argue that it would be better to automatically set the color and size of the ads, and deliver stylewise ads to make the page aesthetic and, thus, reduce the intrusiveness of the ads.[2]
3) Most existing contextual ad-networks require publishers to predefine the position of ads. People tend not to look at ads located at fixed positions in the Web pages [4]–[6]. We believe that it is better to place ads in different positions to avoid ad blindness. Therefore, some publishers change their ad positions frequently to achieve better effect for advertising. If the ad networks can provide publishers an option to automatically select suitable ad positions, it will save a lot of manual work.
4) We believe that adaptive ads are able to reduce the effort of publishers who have a lot of Web pages with different styles. In the AdSense management system, publishers can create and modify ad units. Each ad unit should be predefined according to its color setting. We argue that, when the color and format of ads are predefined in AdSense, they cannot be changed automatically to match the style of the hosting page. Therefore, it is better to have the ad units dynamically change and match the style of the hosting pages when they are placed in different pages.

[2]www.google.com/adsense



(a) (b)

Fig. 3. Ads are inserted in the blank area within the Web page when applying PageSense. (a) Blank area within a Web page. (b) Ads inserted in the blank area.

The above analyses motivate us to design a new mechanism for online contextual advertising. It has been noted that Web pages often have some blank regions, as shown in Fig. 3(a). Most Web pages are dynamically generated or partially created on servers by back-end technologies, such as active server page and java server pages. Web developers or Webmasters only design the main architectural structure. The content of a page is usually selected from the database and other kinds of local or networked data. Unless the Web pages are edited and formatted carefully by a professional, they usually have some unexpected empty space due to the different lengths of source content. Various organizations use Web template systems for mass-production of Web documents. There is no guarantee that there would not be any empty space within such dynamically generated Web pages. Most subpages have blank areas created accidentally.

Although some publishers deliberately design a "relax page" with specific blank areas, the unexpected blank regions created by dynamic Web page technologies will sometimes make users feel a little uncomfortable. Thus, we argue that embedding compelling ads that are visually consistent with the host page into the blank regions can complement the hosting page [7]. We conducted a study based on the sizes that Google's ads use. According to Alexa.com's traffic statistics, we analyzed the homepages of the top 10 000 Websites. We find that nearly 90% of Web pages contain blank areas, which are large enough (larger than 10 000 pixels) for inserting ads. As we have mentioned, current online contextual advertising can neither properly consider style matching nor dynamically deliver stylewise ads in suitable positions on the Web page. In this paper, we attempt to deliver suitable advertisements to the blank regions of the Web page.

Motivated by the above analysis, we propose a novel contextual advertising system, named PageSense. It supports contextual advertising by associating the most relevant ads with the Web page content as well as the Web page style in suitable blank areas within a page. Style refers to a particular form of appearance and design. In our paper, Web page "style" mainly refers to the color and layout of a Web page. We believe that a consistent visual style between ads and hosting Web pages will improve user experience consuming ads. PageSense also aims at releazing Website publishers from changing ad appearance and position manually and frequently. Fig. 3(b) shows an example subscribing PageSense's service. The benefits of delivering style-consistent ads in the blank regions of Web pages are 1) the embedded ads that are consistent with the style of the ad landing page will more likely complement the blank area and improve the user browsing experience; 2) no predefined ad blocks are needed, which saves publishers from having to apply a contextual advertising service; and 3) delivery of style-consistent ads can reduce ad intrusiveness, and thus improve the user ad experience.

This paper makes the following contributions.

1) We propose a new contextual Web advertising system, named PageSense, which utilizes style-relevant matching to improve advertising relevance. Compared with conventional contextual advertising, PageSense is able to choose style-consistent ads dynamically for hosting Web pages to improve user experience.
2) In PageSense, we propose using unexpected blank regions for ad insertion without breaking the original structure of the landing page. It releases the publishers from rearranging the layout of the hosting Web page manually. We propose a new vision-based method for detecting ad insertion positions on the Web pages. Our method is simple yet robust compared with existing methods based on HTML analysis.
3) We conduct a series of in-depth user studies to verify the effectiveness of PageSense as a powerful complement to existing contextual advertising.

The rest of this paper is organized as follows. Section II reviews related works. Section III describes the proposed PageSense system. Section IV presents experiments and evaluations, followed by discussions in Section V.

II. Related Work

Extensive research has been conducted for online advertising, which features three important problems, i.e., ad relevance, placement, and user experience.

A. Ad Relevance Matching

Research on ad relevance matching can be further classified into two categories from the perspective of what the ads are matched against: contextual advertising and user-targeted advertising.

1) Contextual Advertising: Contextual advertising refers to the placement of commercial advertisements within the content of a generic Web page based on similarity between the content of the target page and the ad description provided by the advertiser. Yih *et al.* [8] propose a method for extracting keywords from pages, using a supervised approach and a corpus of pages whose keywords are manually identified. They show that a model using logistic regression outperforms traditional vector models with tf-idf term weights for keyword extraction. Broder *et al.* [2] classify both target pages and ads into an extensive taxonomy, and then match ads to target pages falling in the same node in the taxonomy. Each node in the taxonomy is represented by a set of bidding phrases or queries corresponding to a concept. Xu *et al.* [9] enrich the semantic expression of a target page (or an ad) by using Wikipedia thesaurus knowledge. Ribeiro-Neto *et al.* [10] focus on the vocabulary mismatch problem, noting that the lack of overlap in the vocabulary of the ads and the vocabulary of the target page degrades accuracy. They refer to this as the vocabulary impedance problem. To overcome this limitation, they expand the vector representation of the target page to increase the chance of matching some of the terms in the ad. Wu *et al.* [11] select relevant ads based on both global context relevance and local context relevance.

2) User-Targeted Advertising: In user-targeted advertising, ads are driven based on user profile and demography or behavior. A user's profile and demography often contain the age, gender, education, income, interests, and other information about the user. Using user profiles to match ads is an explicit method. However, user profiles may be incomplete or inaccurate. Because users update their profiles frequently, the profiles may not be up to date. For example, the personalized ad delivery in interactive digital television is a potentially popular application [12]–[14]. Such advertising refers to the delivery of advertisements tailored to the individual viewers' profiles on the basis of knowledge about their preferences [13], current and past contextual information [12], or sponsors' preferences [14]. In these systems, the users are grouped into a set of predefined interest groups, or their profiles are provided in advance, and then ads that most likely will interest the users will be delivered based on text matching and classification techniques.

User activity and behavior can be a better source for understanding users in an implicit way. For instance, a user who likes to listen to 50-s music may be more senior in age. User-generated content, such as blogger/twitter posts and mobility context, are other sources for understanding users [15]. In [16], Provost *et al.* provide a method for finding good audiences for brand advertising by extracting quasi-social networks from browser behavior on user-generated contents. Other methods make ads user-targeted in terms of geolocation [17], [18], social context [19], [20], and so on.

B. Ad Position Detection

In addition to contextual relevance, it is important to insert an ad into a nonintrusive position in the hosting medium. Mei *et al.* [21], [22] examine how to detect appropriate insertion positions in a video stream or within an image to reduce intrusiveness. Contextual relevance deals with the selection of relevant ads according to the given media (image or video), while content intrusiveness answers the question

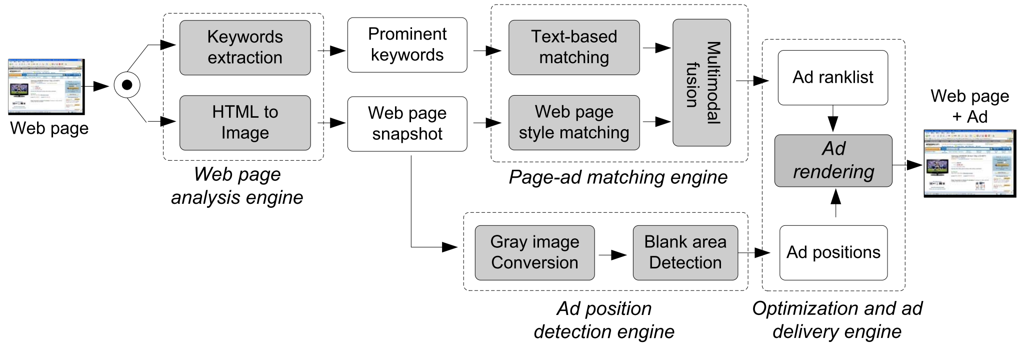


Fig. 4. System framework of PageSense.

of where the selected ads should be embedded. There are various strategies for finding ad insertion points through content intrusiveness. One way is to find the most interesting or highlighting segments in a video or the most salient region in an image as the ad insertion point, so that the ad impression will be maximized [1]. The other is contrary, i.e., to seek the most nonsalient part as an ad insertion point, so that users may not feel intruded upon when browsing the augmented medium with ads. Finding suitable insertion points is actually a kind of tradeoff between ad impression and viewing experience, which deserves a deep study into the effectiveness of advertising [23]. We next describe how to detect ad insertion points in the image and video domain through content intrusiveness analysis.

1) Ad Position Detection in Image Domain: In contextual image advertising, the relevant ads are embedded at certain spatial positions within an image. The image ad-network finds nonintrusive ad positions within an image and selects the ads whose product logos are visually similar or have a similar visual style to the image to minimize intrusiveness and improve user experience. In particular, in [22] and [24], the candidate ad insertion positions (usually image blocks) are detected based on the combination of an image saliency map, face detection, and text detection, while visual similarity is measured on the basis of HSV color features. In this way, a combined saliency map is obtained in which the value of each pixel indicates the overall salience for ad insertion.

2) Ad Position Detection in Video Domain: The ads in contextual video advertising can be inserted in a preserved page block around the video [25], as an overlay video on certain frames [26], in the story or scene breaks in a video [21], [27], [28], or even into a spatiotemporal portion of a video [29]–[32]. The detection of in-stream ad insertion point is based on content intrusiveness [21], [27]. Two computable measurements based on these eight factors, i.e., content discontinuity and attractiveness, are excerpted from video signals. The detection of ad insertion points can be formulated by ranking the shot boundaries based on different combinations of content discontinuity and attractiveness. In contrast to video advertising for general video content, to make video content more enriching, some researchers have attempted to spatially replace a specific region with product advertisements in sports videos [29], [31], [32]. These regions could be locations with less information in a baseball video [31], tennis video [29], or the smooth regions in sports videos [32]. Li *et al.* [30] propose finding the most nonsalient space-time portions of the video. They formulate the problem as a maximum *a posteriori* problem, which maximizes the desired properties related to a less intrusive viewing experience, i.e., informativeness, consistency, visual naturalness, and stability. For a more comprehensive survey, please refer to [23].

C. User Experience on Advertising

A beautifully designed Web page layout will definitely facilitate users' experience when they are browsing the pages. Users have positive attitudes toward Web advertising if the overall Web page looks beautiful [33], [34]. Some researchers have also investigated the cognitive problems in Web advertising, especially using eye tracking as a performance metric for online advertising [35]–[37]. Based on these literatures, we can conclude that Web page aesthetic and ad positions play an important role for advertising effectiveness. However, the way to keep a good balance between advertising intrusiveness to users and the aesthetic appearance of a Web page still remains a challenging research problem.

III. PageSense System

A. PageSense Overview

Fig. 4 shows the framework of PageSense. The system consists of the following components.

1) Web page analysis engine. This contains a keyword extraction module, a Web page categorization module, and an HTML-to-image module. The keyword extraction module extracts prominent keywords from the original Web page. The Web page categorization module classifies Web pages into predefined categories. Each category is represented by some keywords. The HTML-to-image module creates a snapshot of the Web page.
2) Page-ad matching engine. The page-ad matching engine matches the Web page with ads based on category relevance, text relevance, and style consistency. It has three modules: a) The text-based ranking module searches relevant ads based on extracted keywords; b) the page style matching module filters nonrelevant ads using style information; and c) the multimodal fusion module combines the above two modules to achieve optimized results from the ad database.

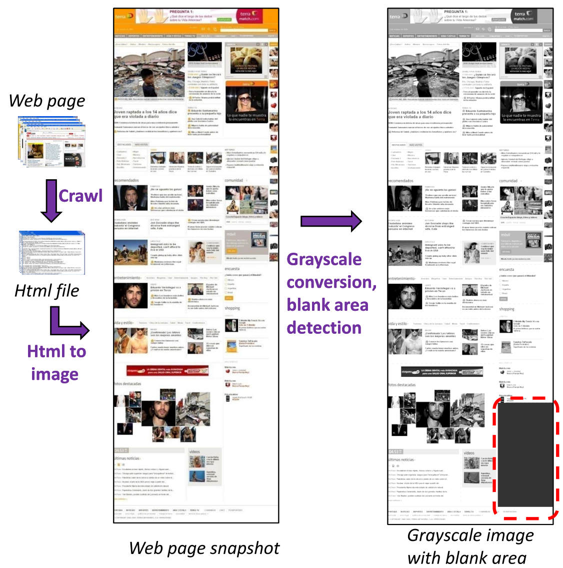


Fig. 5. Ad insertion position detection. The Web page is first crawled and converted into a snapshot image and further transformed into a grayscale image. The highlighted rectangle area indicates the largest blank area that is selected as the ad insertion position.

3) Ad position detection engine. The ad insertion point detection engine converts the Web page snapshot (color image) into a gray image first and then detects blank areas for ad insertion. The informative content of the host page, therefore, will not be occluded, and users will not feel intrusive with the inserted ads.
4) Ad delivery engine. The ad delivery engine delivers and renders the relevant ads in blank areas within a Web page. To build the new Web page with ads, we only need to add one piece of HTML code [more on this in Section IV-E (Fig. 11)] at the bottom of the Web page. After the current Web page is crawled and analyzed, the HTML layers embedded with the ads are delivered to the browser and overlaid onto the blank areas.

To increase flexibility, PageSense further provides two options for publishers. If publishers choose the positions within the Web page for ad placement by themselves, PageSense will omit the position detection for ad placement. If publishers do not specify the ad positions, PageSense will do the complete process, which can automatically detect the most nonintrusive blank region for ad insertion, select the most relevant ads, and deliver them to the blank areas without breaking the original style of the Web page. In Sections III-B–D, we will introduce the key components of PageSense.

B. Web Page Analysis and Ad Position Detection

In order to embed ads at appropriate positions in a Web page, we need to achieve balance between the overwhelming of Web page content and the overlooking of ads from users. The process of ad position detection is shown in Fig. 5. First, the Web page is crawled and converted into a snapshot image. Then, the snapshot image is processed and converted into a grayscale image. We detect the blank areas within the grayscale image based on image analysis. Finally, we calculate the ad position relative to the original page. Blank areas are the candidate positions for delivering an ad.

We use the hole detective algorithm to detect the blank area within a Web page. The blank area is the candidate position for delivering an ad. Algorithm 1 shows the process for blank area detection. Let p_{ij} denote the grayscale of 8 b of each pixel (x_i, y_j). N_h denotes the number of holes $H = \{H_k\}_{k=1}^{N_h}$, while S denotes the size of the hole H_k. Let h_{ij} denotes which hole the pixel (x_i, y_j) belongs to $h_{ij} \in \{0, 1, 2, \ldots, N_h\}$. When $h_{ij} = 0$, it means pixel (x_i, y_j) does not belong to any holes. Essentially, each pixel can be treated as a very small hole. However, if the size of the hole is less than a predefined threshold e, we can discard it. If the pixel (x_i, y_j) has not been processed, we set $h_{ij} < 0$.

In Algorithm 1, we initially set $h_{ij} = p_{ij} - 256$. Because $p_{ij} \leq 255$, we have $h_{ij} < 0$. It means that not every pixel (x_i, y_j) has been processed yet. In Step 3, we go through all pixels of the image. Finally, we get the number of holes within the image and the sizes of these holes. Since each pixel (x_i, y_j) is labeled with the hole to which it belongs, i.e., h_{ij}, it is easy to get the position of the holes. In PageSense, we use the top-left pixel and the width/height attributes to mark the holes, i.e., the blank area for ad insertion. Intuitively, the more ad positions there are, the more intrusiveness will be caused. In PageSense, we usually select the largest blank area as the only ad position.

C. Page-Ad Matching for Candidate Ad Ranking

As we have mentioned, to select relevant ads according to the hosting Web page, we need to consider textual relevance between the ad description and the Web page content, the visual style (i.e., color and layout) relevance between the ad appearance and the Web page, as well as the local visual similarity between the ad appearance and the neighboring background of the Web page. In the next part, we describe how we can compute these types of relevance.

1) Text Ranking: PageSense supports more effective contextual Web advertising in terms of text relevance, style relevance, and intrusiveness. Therefore, PageSense should be able to select semantically relevant ads. The traditional vector space model is far from a semantic requirement for ad selection in contextual advertising [38]. Instead, topic models are widely used, since the models are able to consider the correlation between terms and generate a topical representation of the inherent structure of the corpus. Therefore, they are more reasonable than exact term representation and widely applied in many retrieval tasks. Among topic-based language models, latent Dirichlet allocation (LDA) is well known for its good ability to discover the latent topic structure [39], [40]. Therefore, we adopt LDA to compute semantic relevance between ads and Web pages in PageSense.

LDA is a generative probabilistic model of a corpus. Let a word w be a unit-basis vector with the ith component equal to one and all other components equal to zeros, if it is the ith word in the vocabulary. Given a corpus D containing V unique words and M documents, D are represented as random mixtures over a set of n latent topics $\mathbf{z}$, where each topic z is

Algorithm 1 Hole Detection

Input: $p_{i,j}$, which is a gray scale of each pixel, $i = 1, \ldots, H,\ j = 1, \ldots, W$
Output: $h_{i,j}, k, S$
Initialization: set $h_{i,j} = p_{i,j} - 256, k = 1, S = 0$
1: **while true do**
2:   **if** there is $h_{i,j} \leq 0$ **do**
      set $h_{i,j} = 0$; go to step 3
    end if
3:   **while true do**
      for $i = 1 : H$ **do**
        for $j = 1 : W$ **do**
          if $h_{i,j} \leq 0$, **do**: continue this iteration;
          end if
          if $h_{i,j-1} == k$ or $h_{i-1,j} == k$ **do**
           $h_{i,j} == k$;
          end if
        end for
      end for
      for $i = H : 1$ **do**
        for $j = W : 1$ **do**
          if $h_{i,j} \leq 0$, **do**: continue this iteration;
          end if
          if $h_{i,j+1} == k$ or $h_{i+1,j} == k$ **do**
           $h_{i,j} == k$;
          end if
        end for
      end for
      if there is no h_{ij} set to k **do** end the iteration
      end if
4:   **end while**
5:   Generate $T = 0$
    for each iteration i, j **do**
      if $h_{i,j} == k$, **do**: $T++$;
      end if
    end for
6:   **if** $T < e$ (e is a threshold) **do**
      for each iteration i, j **do**
        if $h_{i,j} == k$, **do**: $h_{i,j} == 0$;
        end if
      end for
    else $S+ = T, k++$;
    end if
7: **end while**; $k--$
8: **Return** $h_{i,j}, k, S$

characterized by a distribution over words. The LDA model defines two corpus-level parameters α and β, where α is a k-vector of Dirichlet parameters, β is a $K \times V$ matrix of word probabilities, and $\beta_{ij} = p(w_j = 1|z_i = 1)$, where $i \in \{0, 1, \ldots, K\}$ and $j \in \{0, 1, \ldots, V\}$.

We consider Web page d_p as the query and ad d_a as the document. Each ad can be represented by $d_a = (d_{a1}, d_{a2}, \ldots, d_{an})$. The LDA model regards the given query $d_p = (d_{p1}, d_{p2}, \ldots, d_{pm})$ as an unseen document. Given an LDA model (α, β) and an advertisement d_a, the similarity between d_p and d_a is defined by

$$\mathrm{sim}(d_p, d_a) = p(d_p|d_a) = \prod_{i=1}^{m} p(d_{pi}|d_a, \theta, \beta). \tag{1}$$

The inference process of $p(d_{pi}|d_a, \theta, \beta)$ is defined as

$$p(d_{pi}|d_a, \theta, \beta) = \int p(\theta|d_a, \alpha) p(d_{pi}|\theta, \beta) d\theta \tag{2}$$

where

$$p(\theta|d_a, \alpha) = \frac{p(\theta, d_a|\alpha, \beta)}{p(d_a|\alpha, \beta)} = \frac{p(\theta|\alpha) \prod_{j=1}^{N} p(w_j|\theta, \beta)}{p(d_a|\alpha, \beta)}. \tag{3}$$

Based on (1)–(3), we can calculate the text relevance between the Web page and the ad. Let $W^{(T)}$ and $a_j^{(T)}$ denote the textual information of Web page W and the ad a_j; the textual relevance $R_t(W, a_j)$ is given by

$$R_t(W, a_j) = \mathrm{sim}\big(W^{(T)}, a_j^{(T)}\big). \tag{4}$$

In summary, the Web pages are classified into eight predefined categorizations: 1) Business and Economy; 2) Computer and Internet Info; 3) Educational Institutions; 4) Entertainment and Arts; 5) Fashion and Beauty; 6) Financial; 7) Travel; and 8) Sports. We leverage the WordNet-based keyword expansion technology for the names of the categories to get more concepts/keywords. These keywords and the prominent keywords extracted from the Web pages are then used as queries to retrieve relevant ads.

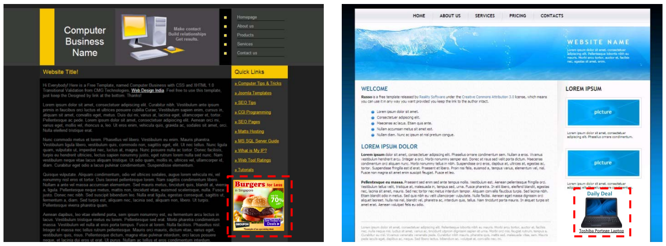
2) Style Ranking With Visual Analysis: Intuitively, the ads are expected to have a similar style as the hosting Web page, so that users may perceive the ads as a natural part of the original Web page. We use global color features for measuring appearance similarity. Let $W^{(V)}$ and ${a_j}^V$ denote the global visual information of Web page W and ad a_j, respectively. To measure the similarity between a_j and W, we compute the L_1 distance between $a_j^{(V)}$ and the Web page W^V. The distance is defined in the HSV color space, which has been widely adopted in multimedia information retrieval

$$d\big(W^V, a_j^{(V)}\big) = \sum_{k=1}^{K} |f^{(w)}(k) - f^{(a_j)}(k)| \tag{5}$$

where $f^{(w)}(k)$ and $f^{(a_j)}(k)$ denote the kth color feature of Web page W and ad a_j, respectively. K is the feature dimension. As a result, the local content relevance $R_s(W, a_j)$ between a Web page W and ad a_j is given by

$$R_s(W, a_j) = 1 - d\big(W^V, a_j^{(V)}\big). \tag{6}$$

A PageSense ad is made up of the ad title, the ad body copy (images and text description), and the destination URLs. In order to maintain a consistent style after the ads are inserted, PageSense will use colors for the ad text and links that already exist on the hosting Web page by a color extraction tool for Web pages. For example, if the links on the Web page are all green and the text is black, PageSense uses green links and black text in the embedded ads as well. Since most users are accustomed to seeing blue links, if PageSense fails to extract the Web page color, it will use blue as the default setting. In this way, PageSense can dynamically change the style of the inserted ad based on the style of the hosting Web page.



(a) (b)

Fig. 6. Strategies for local visual matching. (a) Contrasting. (b) Blending.

3) Local Visual Ranking: In order to make embedded ads look natural while minimizing visual intrusiveness, the ads should be inserted into nonintrusive blank areas and be visually relevant to the background color of the insertion areas. Based on image processing methods, PageSense can easily detect the background color of any regions from the snapshot image of the Web page. Therefore, we have the following strategies for local visual matching, depending on the color of the background where the ads are placed.

1) *Contrasting:* For Web pages with a dark background, PageSense will select the ads with colors that stand out against the background, for example, one with a light yellow background and light blue titles, as shown in Fig. 6.
2) *Blending:* For Web pages with a light background, PageSense makes the background and borders of the ads the same color as the background of the region where the ad is placed. For example, if the Website has a white background, PageSense will choose ads with a white background, as shown in Fig. 6.

D. Optimization-Based Ad Delivery

Given candidate ads and ad insertion areas obtained in previous sections, our objective is to associate the appropriate ad with each candidate location. We adopt an optimization-based approach to maximize overall advertising relevance.

1) Problem Formulation: Let W denote the hosting Web page, which has N_p candidate ad insertion regions that are represented by $\mathcal{P} = \{p_i\}_{i=1}^{N_p}$. The problem with online ad insertion can be described as given a set of insertion points $\mathcal{P}$ and a list of ranked ads $\mathcal{A}$ to select N elements from $\mathcal{P}$ and $\mathcal{A}$, respectively, and to associate each $a_j \in \mathcal{A}$ with an appropriate $p_i \in \mathcal{P}$. N is the number of ads expected to be inserted. To support contextually relevant and less-intrusive advertising from the perspective of viewers, several computable objectives can be expressed as follows.

1) *Text Relevance $R_t(W, a_j)$:* It measures the relevance between keywords extracted from Web page W and the text information associated with the ad a_j. We use an LDA-based topic model to compute text relevance between ads and Web pages, as shown in (4).
2) *Global Style Relevance $R_s(W, a_j)$:* It measures the style consistency between ad a_j and page W. The product logo or thumbnail is the main component of ad a_j in PageSense. These images are assumed to have an appearance that is similar to the page so that users will perceive the ad as a natural part of the original page. To measure the visual similarity between a_j and W, we compute the $L1$ distance defined in the HSV color space by (6).
3) *Local Visual Relevance $R_l(W, a_j)$:* It measures the visual relevance between the ad insertion position b_i and the ad a_j, as well as the nonintrusiveness of b_i for inserting ads. To minimize intrusiveness to the user, the ads are to be inserted into nonintrusive blank positions and be locally relevant to the background of the positions, as described in Section III-C3.

Suppose we introduce the following design variables $\mathbf{x} \in \mathbb{R}^{N_p}$, $\mathbf{y} \in \mathbb{R}^{N_a}$, $\mathbf{x} = [x_1, \ldots, x_{N_p}]^T$, $x_i \in \{0, 1\}$, and $\mathbf{y} = [y_1, \ldots, y_{N_p}]^T$, $y_j \in \{0, 1\}$, where x_i and y_j indicate whether p_i and a_j are selected ($x_i = 1, y_j = 1$). The above problem can be formulated as the following nonlinear 0-1 integer programming problem:

$$\begin{aligned}\max_{(\mathbf{x},\mathbf{y})}\; f(\mathbf{x}, \mathbf{y}) &= w_t \sum_{j=1}^{N_a} y_j R_t(a_j) + w_s \sum_{j=1}^{N_a} y_j R_s(a_j)\\ &\quad + w_\ell \sum_{i=1}^{N_b}\sum_{j=1}^{N_a} x_i y_j R_\ell(b_i, a_j)\\ &= w_t \mathbf{y}^T \mathbf{R}_t + w_s \mathbf{y}^T \mathbf{R}_s + w_\ell \mathbf{x}^T \mathbf{R}_\ell \mathbf{y}\\ \text{s.t. } \sum_{i=1}^{N_b} x_i = N,&\quad \sum_{j=1}^{N_a} y_j = N, \quad x_i, y_j \in \{0, 1\} \end{aligned} \tag{7}$$

where $\mathbf{R}_t = [R_t(a_1), R_t(a_2), \ldots, R_t(a_{N_a})]^T$, $\mathbf{R}_\ell = [R_\ell(a_1), R_\ell(a_2), \ldots, R_\ell(a_{N_a})]^T$, and $\mathbf{R}_\ell \in \mathbb{R}^{N_b \times N_a}$. The parameters (w_t, w_s, w_ℓ) control the emphasis on global and local textual relevance, as well as local content relevance, and satisfy the constraints: $0 \leqslant w_t, w_s, w_\ell \leqslant 1$ and $w_t + w_s + w_\ell = 1$. The parameters can be trained by cross validations. Since our system supports multiple ads on the same Web page, the N ads $\mathcal{X} = \{x_i\}_{i=1}^{N}$ with the highest equation scores in (7) are selected as relevant ads that will be delivered to the Web page. N is decided by the publisher and the default value is 5.

It is worth emphasizing that (7) is mainly designed from the perspective of publishers and consumers. However, this formulation can be easily revised for advertisers by adjusting the weights of both local visual relevance and global style relevance in (7). In this way, we can make the ads look much more salient in terms of visual contrast.

2) Problem Solution: There are $C_{N_a}^{N} C_{N_b}^{N} N!$ solutions in total to (7). When the numbers of elements in $\mathcal{A}$ and $\mathcal{B}$ are large, the searching space for optimization is very large. Therefore, we use the genetic algorithm (GA) to solve this problem efficiently. GAs are implemented in a computer simulation in which a population of genetic representations (called chromosomes) of candidate solutions (called individuals) to the optimization problem evolves toward better solutions. The evolution usually starts from a population of randomly generated individuals and happens in generations. In each generation, the fitness of every individual in the population is evaluated, and multiple individuals are stochastically selected

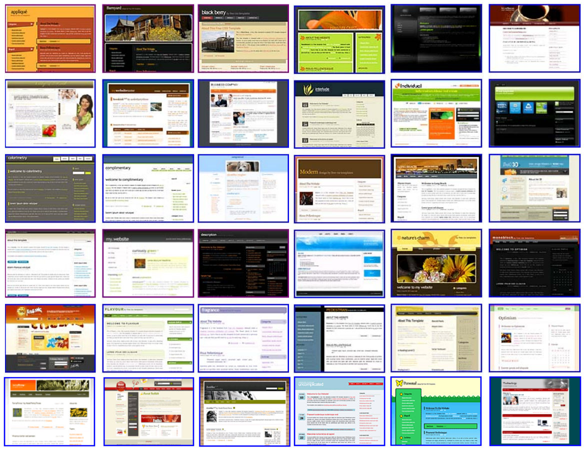


Fig. 7. Web page examples.

TABLE I
Ad Product Types

	Product categories	Number	Percent
1	Books	1094	9.13%
2	Movies, Music & Games,	773	6.45%
3	Digital Downloads	1019	8.5%
4	Kindle	3	0.025%
5	Computers & Office	897	7.48%
6	Electronics	1241	10.35%
7	Home & Garden & Pets	1062	8.86%
8	Grocery, Health & Beauty	884	7.38%
9	Toys, Kids & Baby	956	7.98%
10	Clothing, Shoes & Jewelry	981	8.19%
11	Sports & Outdoors	1193	9.95%
12	Tools, Home Improvement	975	8.16%
13	Automotive & Industrial	907	7.57%

from the current population based on their fitness and randomly mutated to form a new population. The new population is then used in the next iteration of the algorithm. Algorithm 2 gives the algorithm. The algorithm terminates when either a maximum number of generations has been produced, or a satisfactory fitness level has been reached for the population.

IV. Experiments

In this section, we systematically evaluate PageSense, including relevance, user experience, and usability.

A. Data Set

Based on the Web Directory on Alexa.com, which lists the top sites in each category, we chose one Web page from each category. In total, we collected 672 pages for evaluation. Since these Web pages were selected from different categories, they covered a wide range of styles. The example Web page collection data are shown in Fig. 7. These Web pages are downloaded and deployed on our own server, and therefore, PageSense ads and AdSense ads can both be inserted for comparison.

We also collected Amazon products with 13 categories. There are a total of 11 895 ads, which are illustrated in Table I. Each advertisement includes product title, Amazon Standard Identification Number, model number, average customer review score (0–5), sales rank, description, detail page URL, price, categories, and logos.

Algorithm 2 Genetic Algorithm for (7)

1: Set initial and global parameters.
Set population size N_p, probability of crossover P_c, the probability of mutation P_m and maximum generation $maxgen$. Set the initial generation $gen = 0$, the initial maximum fitness $maxeval = 0$.

2: Generate the initial chromosome υ_k $(k = 1, \ldots, N_p)$ with t genes randomly satisfying the following conditions: $\sum_i^{N_b} x_i = N$, $\sum_j^{N_a} y_j = N$, $x_i, y_j \in \{0, 1\}$. The calculations stop if we cannot find a chromosome satisfying the above conditions.

3: Calculate the fitness value.
Set $gen = gen + 1$, and calculate evaluation function $eval(\upsilon_k)$ for each chromosome, $eval(\upsilon_k) = w_g R_g(\upsilon_k) + w_\ell R_\ell(\upsilon_k) + w_c R_c(\upsilon_k), k = 1, 2, \ldots, N_p$.

4: Genetic operation.

4.1 Crossover:
(1) Let the number of chromosomes generated by crossover be $N_c = 0$.
(2) Create random number $r_k \in [0, 1]$ $(k = 1, 2, \ldots, N_p)$.
(3) Select υ_k that satisfies $r_k < P_c$. Make a pair of υ_k and set $N_c = N_c + 2$.
(4) Choose the position for crossover at random and undergo crossover. Let the offsprings that are newly generated be υ'_{N_c-1} and υ'_{N_c}, where $\upsilon'_{N_c-1} = \lfloor c\upsilon_j + (1-c)\upsilon_l \rfloor$, $\upsilon'_{N_c} = \lfloor c\upsilon_l + (1-c)\upsilon_j \rfloor$, $(j, l \in \{1, \ldots, N_p\}, 0 \le c \le 1)$, $\lfloor x \rfloor$ is defined as the maximum integer smaller than real number x, and $c = \frac{eval(\upsilon_k)_{\max}}{eval(\upsilon_k)_{\min} + eval(\upsilon_k)_{\max}}$, where $eval(\upsilon_k)_{\max}$ and $eval(\upsilon_k)_{\min}$ are the maximum and minimum fitness of chromosomes in generation gen.

4.2 Mutation:
(1) Let the number of chromosomes generated by mutation be $N_m = 0$.
(2) Create random number r_k $(k = 1, \ldots, n \times N_p + n \times N_c)$ from [0, 1].
(3) Select a gene that satisfies $r_k < P_m$ and set $N_m = N_m + 1$, and proceed with local search-based mutation.
(4) Let the newly generated chromosome be $\upsilon'_{N_c+N_m}$.

4.3 Selection:
(1) Calculate the evaluation function $eval(\upsilon'_t)$, where $t = 1, \ldots, (N_c + N_m)$.
(2) Select the chromosomes from among the parents $\{\upsilon_k | k = 1, \ldots, N_p\}$.
(3) Set the newly generated offspring $\{\upsilon'_t | t = 1, \ldots, N_c + N_m\}$ in the way that is superior to others.
(4) Duplicate selection is prohibited. The number to be selected is N_p.
(5) Let the chromosomes $\{\upsilon'_k | k = 1, \ldots, N_p\}$ enter the next generation.

5: If $maxeval < max\{eval(\upsilon'_k) | k = 1, \ldots, N_p\}$, then
$maxeval = max\{eval(\upsilon'_k) | k = 1, \ldots, N_p\}$,
$\upsilon^* = argmax_{\upsilon_k}\{eval(\upsilon'_k) | k = 1, \ldots, N_p\}$.

6: If $gen < maxgen$ then goto Step 3. If $gen = maxgen$, then output υ^* and stop.

B. Evaluation of Ad Relevance

In order to compare PageSense with AdSense, we joined the AdSense program as a publisher, and added AdSense ads to those Web pages. Considering that PageSense mainly uses image-based ads, we set AdSense to only deliver image ads though the configuration tool of Google AdSense.

Each page-ad pair was judged by three or more subjects on a (1–3) scale as follows.

TABLE II

EVALUATIONS ON ad RELEVANCE

	AdSense	PageSense
Average Value	2.75	2.86
Accuracy	89%	92%

1) *Irrelevant (1):* The ad is definitely unrelated to the hosting Web page.
2) *Somewhat Relevant (2):* The ad is related to the secondary subject of the Web page, or related to the Web page in a general way.
3) *Relevant (3):* The ad is directly related to the main subject of the Web page content.

In particular, each page-ad pair was first displayed for three subjects. If all three subjects or at least two gave the same score, this score was taken as the final judge result for this page-ad pair. Otherwise, two more subjects were involved to give scores for this page-ad pair. In this case, a final score was taken if at least three out of five subjects agreed with each other on this score. Otherwise, this sample would be discarded.

Table II lists the evaluation results. We observe that PageSense provides better ad relevance. The average relevance score of PageSense is 2.86, while AdSense achieves a score of 2.75 for ad relevance. The standard deviation of the PageSense ad relevance score is 0.46, while that of AdSense is 0.82. The results also show the ad relevance in PageSense is statistically superior to AdSense. We assume that the page-ad pairs judged with score of "2" or "3" are positive and those judged with a score of "1" are negative. Based on the number of true positives and the sum of all the examples, we calculate the accuracy of Advertising. We observe that the accuracy of PageSense is higher than AdSense. Here, we only take 1 as negative and an ad with a score of "2" is somewhat relevant to the Web page, although not perfectly relevant. We provide option "2" for subjects, because there are some cases when the ad neither perfectly matches the Web page nor is completely unrelated to it. After we check the scores received from subjects, we find that in most cases, the scores are either "1" or "3." There are very few "2"s. If we take both "1" and "2" as negative, the Accuracy in Table II are 86% (AdSense) and 91% (PageSense).

Since PageSense selects the top K relevant ads for displaying in turn, it is an information retrieval scenario. In information retrieval, precision and recall are usually used to evaluate the performance of the proposed system. A good precision score means that a high percentage of the results retrieved by a search are relevant (but says nothing about whether all relevant documents are retrieved), whereas a good recall score means that many relevant documents are retrieved by the search (but says nothing about how many irrelevant documents are also retrieved). Usually, precision and recall scores are not discussed in isolation. Therefore, we use 11-point precision-recall figures to compare the values of precision for a fixed level at recall. Since AdSense only searches one image ad for an ad unit (we only set one ad unit within the page), whereas PageSense is able to deliver multiple ads in turn for display on the blank area (i.e., ad position) within the page,

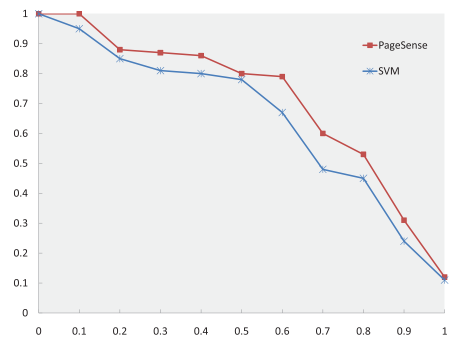


Fig. 8. Precision–recall curve.

we cannot compare PageSense with AdSense in the precision–recall evaluation. In order to evaluate the PR measure of PageSense, we use the support vector machine (SVM)-based ad classification as a baseline for comparison. In the SVM-based method, we first define eight categories of ads and label some training data, and then train corresponding ad classifiers based on the text features of the Web page. Fig. 8 shows the precision–recall curve of PageSense. PageSense achieves a better performance than SVM-base ad classification in terms of relevance. Experiments confirm that LDA model-based PageSense is suitable for advertising. We believe that an optimal set of parameters for the equation can further improve advertising performance.

C. Evaluation of User Experience

We conducted a user study to evaluate the user experience of PageSense; 200 Web pages were randomly selected for evaluation from our 672 Web page data set as mentioned above. Each Web page was subscribed to the AdSense service and PageSense, and 50 subjects, consisting of university students, teachers, medical personnel, IT workers, and finance workers, were asked to compare PageSense with AdSense. There were 25 female and 25 male participants, with the age ranging from 18 to 45. About 16% of them had experience on using contextual ad-networks, such as Google AdSense, Yahoo! Publisher Network, and Bing Ads. The rest of them heard about online advertising and were aware of these ads on some Web pages. The human subjects were asked to give scores (1–5) according to the following questionnaire.

1) Position: How do you feel about the ad position? (5: very good; 4: good; 3: ok; 2: not good; 1: bad.)
2) Impression: Does the ad impress you? (5: very impressive; 4: impressive; 3: ok; 2: not too impressive; 1: unimpressive.)
3) Acceptance: If you owned a Web page, would you be willing to subscribe to PageSense? (5: very willing; 4: willing; 3: maybe; 2: not very willing; 1: unwilling.)
4) Overall effectiveness: How do you think of the overall advertising effects? (5: very good; 4: good; 3: ok; 2: not good; 1: bad.)

Fig. 9 shows the evaluation results. A higher score indicates higher satisfaction. We observe that PageSense achieves better satisfaction than AdSense with respect to all considerations, i.e., position, impression, acceptance, and overall effectiveness.

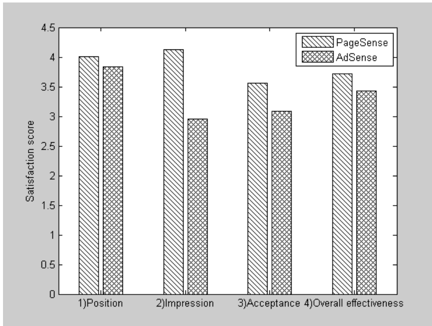


Fig. 9. Evaluations of ad satisfaction.

Although AdSense's ad positions are defined manually and PageSense's are detected automatically, PageSense outperforms AdSense within a slight margin in terms of position. PageSense delivers image-based ads that support a variety of rendering effects. Moreover, PageSense ads are consistent with the landing page's color style. Most of the evaluators found PageSense ads more impressive than AdSense ads. The overall effectiveness score of PageSense is better than that of AdSense, even though the former does not use manual insertion and settings. The evaluation results have proved that the advertising approach of PageSense is effective.

Fig. 10 further shows an example of Web page subscribing PageSense service. We can see that there is a blank region in this Web page, which makes the Web page unattractive. PageSense is able to deliver relevant and style-consistent ads into the blank regions. Fig. 10 shows an example of Web page with relevant overlay ads. We can see that 1) some associated ads are inserted into the empty region of the original Web page, 2) the color of the ads matches the hosting Web page's color, 3) the inserted ads make the Web page more attractive, and 4) the inserted ads are contextually relevant to the ad landing page.

D. Evaluation of Ad Impression

We invited the participants in former user study to evaluate the advertisement impressions. Each of them was asked to join the user study without discussion, and without even knowing the intention of the user study in advance. Each subject was asked to open a Web page with PageSense ads or AdSense ads, and look at it for 10 s, putting himself or herself completely in the mindset of a regular Web page browsing user. We then gave the subject a questionnaire with 30 advertisements and asked him or her to identify which, among the candidate set, had been browsed. The recall rates are shown in Table III. We can conclude that PageSense achieves higher recall than AdSense.

E. Evaluation of Efficiency

AdSense requires a longer time to rearrange the Web page, define the color, and other style attributes of the ad, as well as insert the ad into the page. On the other hand, PageSense only needs publishers to add one piece of HTML code at the bottom of their own Web pages. The inserted HTML code, shown in Fig. 11, is able to download a javascript file that handles ad delivery. First, the current page's URL is sent back to a Web service in the back-end. Second, the Web service will crawl the Web page and detect the blank areas. Third, the HTML layers embedded with ads will be delivered to the browser and overlay on the blank areas.

TABLE III

EVALUATIONS ON RECALL RATE AND SETUP TIME

	AdSense	PageSense
Recall rate	49.13%	61.45%
Setup time	352 seconds	21 seconds

TABLE IV

EVALUATION ON ads USING PREDEFINED ad POSITIONS

Method	Aesthetics	Impression	Intrusiveness	Overall
Original	3.31	3.21	3.19	3.59
Random	2.81	2.97	2.44	2.59
PageSense	3.37	3.98	3.43	3.63

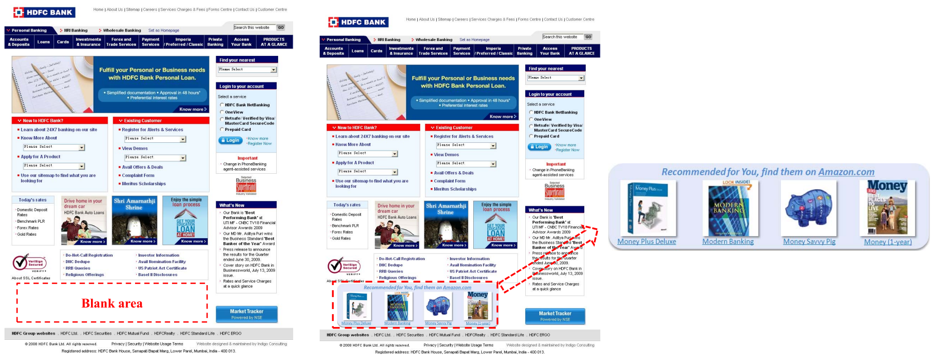
We evaluate the average setup time for AdSense and PageSense. The results are shown in Table III. We conclude that PageSense is easier to use than AdSense. PageSense can significantly save time in setting up an advertising platform. We also evaluate the absolute run time for the complete process of PageSense, including downloading the Web page, analyzing the page content, detecting the ad position, and selecting the proper ads from database, which contains 11 895 ads. The average time for all test Web pages is 0.752 s. About 80% of the total processing time is spent on downloading the Web pages. The PageSense system is viable in practice.

F. Evaluation on Ads Using Predefined Ad Positions

As stated previously another option for PageSense is enabling publishers predefine ad positions in their Web pages manually. We make use of 200 real Web pages, which have had ads inserted. We manually remove the original ads from the pages and empty regions are reserved for ad insertion. We run PageSense to select the optimal ads for the empty regions. To evaluate ad selection, we invite 50 users to score these Web pages with inserted ads. Each Web page has three kinds of different ads embedded at same location: 1) original ads; 2) randomly generated ads; and 3) ads selected by PageSense. The subjects were asked to score the Web pages from the aspects listed below, with 1 the lowest and 5 the highest score.

1) Aesthetics: Do you find the Web page layout is aesthetically pleasing?
2) Impression: Does the advertisement impress you?
3) Intrusiveness: Is the inserted advertisement intrusive to your browsing?
4) Overall: How do you rate the effectiveness of the advertisement?

The results of user evaluation are listed in Table IV. It is obvious that PageSense offers better advertisement selection results than random selection, and comparable results with the original manually selected ads. We also compare PageSense's automatic selection for ad position with manual selection.



(a) (b)

Fig. 10. Advertising example using PageSense. (a) Blank area within the Web page (the highlighted rectangular). (b) Some relevant ads are inserted into the blank area. The highlighted rectangle indicates the inserted ads that are relevant and style consistent with the ad landing page.

```
<script type="text/javascript"><!--
ad_client = "abc-0102034455689124";
ad_slot = "a45r3-0934";
//-->
</script>
<script type="text/javascript"
src = "http://ad3.adservice.com/ad/showads.js">
</script>
```

Fig. 11. Piece of code for PageSense.

In the above Web page data set, the original ads are manually removed and the ad regions are reserved for evaluation. PageSense achieves 85% of the same ad positions with the ones by manual definition. There are 92% users who are satisfied with ad positions automatically selected by PageSense.

V. Conclusion and Future Work

In this paper, we present PageSense—a novel contextual advertising platform supporting not only semantically relevant advertising but style-consistent advertising as well. This platform is able to automatically match ads to the content and style of a Web page, delivering relevant ads in a suitable blank area without disrupting the original structure of the Web page.

We have run a half-year trial on the different formats and placement of AdSense ads and concluded as follows.

1) If we want the biggest revenue impact for the smallest effort, it is important to optimize the ads' color palettes. Defining the right color palettes means the difference between ads the users will notice and ads they will skip right over.
2) In AdSense, even if the ads are designed perfectly, it might not work well when the Website audience is composed mainly of repeat visitors. Because the Website's visitors come back day after day, they will probably become blinded to the position of the ads over time regardless of the ad colors. We believe that it is better to use variations of the ad size and ad placement on different pages of the same site.
3) If we put an ad banner in the same spot consistently throughout the site, the CTR is lower than when using different ad placements. It further illustrates that we should change the ad position dynamically to decrease the tendency for users to ignore anything that is separated from the main content of the Website.

Therefore, PageSense is able to not only choose the right color and position but rotate colors or occasionally switch the location of the ads on the page automatically as well. This can avoid ad blindness while avoiding the necessity of AdSense users placing the ads manually. In addition to the above strategies to decrease ad blindness, PageSense can deliver a group of ads in the same region. There are usually three advertisements in the ad group. PageSense rotates the ads within a single page load for a while (e.g., 15 s) depending on the ad rotation setting. The goal of ad rotation is to keep advertising "fresh." If the ad never changes, users are more likely to ignore it.

In this paper, our formulation of PageSense is a generic framework for online advertising, as introduced in (7) in Section III-D. Although we mainly focus on satisfying publishers in this paper, the framework can easily be extended for different perspectives. For example, if we want to satisfy advertisers while sacrificing consumers viewing experiences, we can change the positive weights of both local visual relevance "$+w_l$" and global style relevance "$+w_s$" in (7) to be negative ("$-w_l$" and "$-w_s$"). By doing this, we can always make the ads look much more salient in terms of visual contrast.

There are a number of possible improvements for PageSense. First, although we considered two aspects of style, i.e., color and layout, we would like to improve the effect of stylewise advertising. Different genres of Web pages should have different ad strategies. Deciding how to identify the type of Web page and how to extract the Web template of a Website may be helpful and complementary ways to improve the style matching between ads and Web pages. Second, it still requires additional user studies of a typical viewer to determine the most effective insertion points. For Website publishers, there is a constant battle between money and satisfied visitors. Possible solutions may include: 1) letting publishers decide whether the detected empty areas are suitable for inserting ads or not (for keeping the Web pages' appearance) and 2) employing aesthetic modeling approaches to verify the Web pages' appearance as a postprocess after the ads are embedded [41], [42]. Furthermore, how we can simultaneously take both the viewers and advertisers into consideration still remains an open problem. Third, how to make an ad impressive while maintaining style consistency may also be worth considering.

REFERENCES

[1] S. McCoy, A. Everard, P. Polak, and D. F. Galletta, "The effects of online advertising," *Commun. ACM*, vol. 50, no. 3, pp. 84–88, 2007.

[2] A. Broder, M. Fontoura, V. Josifovski, and L. Riedel, "A semantic approach to contextual advertising," in *Proc. SIGIR*, 2007, pp. 559–566.

[3] X.-S. Hua, T. Mei, and A. Hanjalic, Eds., *Online Multimedia Advertising: Techniques and Technologies*. Hershey, PA, USA: IGI Global, 2010.

[4] W. Albert, "Do Web users actually look at ads? A case study of banner ads and eye-tracking technology," in *Proc. 11th Annu. Conf. Usability Prof. Assoc.*, 2002, pp. 1–8.

[5] L. Granka, H. Hembrooke, and G. Gay, "Location location location: Viewing patterns on WWW pages," in *Proc. Symp. Eye Tracking Res. Appl.*, 2006, p. 43.

[6] J. W. Owens, B. S. Chaparro, and E. M. Palmer, "Text advertising blindness: The new banner blindness?" *J. Usability Stud.*, vol. 6, no. 3, pp. 172–197, 2011.

[7] D. Cohen-Or, O. Sorkine, R. Gal, T. Leyvand, and Y.-Q. Xu, "Color harmonization," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 624–630, 2006.

[8] W.-T. Yih, J. Goodman, and V. R. Carvalho, "Finding advertising keywords on Web pages," in *Proc. WWW*, 2006, pp. 213–222.

[9] G. Xu, Z. Wu, G. Li, and E. Chen, "Improving contextual advertising matching by using Wikipedia thesaurus knowledge," *Knowl. Inf. Syst.*, vol. 43, no. 3, pp. 599–631, 2015.

[10] B. Ribeiro-Neto, M. Cristo, P. B. Golgher, and E. Silva de Moura, "Impedance coupling in content-targeted advertising," in *Proc. SIGIR*, 2005, pp. 496–503.

[11] Z. Wu, G. Xu, C. Lu, E. Chen, Y. Zhang, and H. Zhang, "Position-wise contextual advertising: Placing relevant ads at appropriate positions of a Web page," *Neurocomput.*, vol. 120, pp. 524–535, Nov. 2013.

[12] G. K. Kastidou and R. Cohen, "An approach for delivering personalized ads in interactive TV customized to both users and advertisers," in *Proc. Eur. Conf. Interact. Television*, 2006, pp. 201–208.

[13] G. Lekakos, D. Papakiriakopoulos, and K. Chorianopoulos, "An integrated approach to interactive and personalized TV advertising," in *Proc. Workshop Personalization Future TV*, 2001, pp. 17–25.

[14] S. Velusamy, L. Gopal, S. Bhatnagar, and S. Varadarajan, "An efficient ad recommendation system for TV programs," *Multimedia Syst.*, vol. 14, no. 2, pp. 73–87, 2008.

[15] J. Zhuang, T. Mei, S. C. H. Hoi, Y.-Q. Xu, and S. Li, "When recommendation meets mobile: Contextual and personalized recommendation on the go," in *Proc. ACM Int. Conf. Ubiquitous Comput.*, Beijing, China, Sep. 2011, pp. 153–162.

[16] F. Provost, B. Dalessandro, R. Hook, X. Zhang, and A. Murray, "Audience selection for on-line brand advertising: Privacy-friendly social network targeting," in *Proc. ACM SIGKDD*, 2009, pp. 707–716.

[17] L. Kennedy, M. Naaman, S. Ahern, R. Nair, and T. Rattenbury, "How Flickr helps us make sense of the world: Context and content in community-contributed media collections," in *Proc. ACM Multimedia*, 2007, pp. 631–640.

[18] J. Li, X. Qian, Y. Y. Tang, L. Yang, and T. Mei, "GPS estimation for places of interest from social users' uploaded photos," *IEEE Trans. Multimedia*, vol. 15, no. 8, pp. 2058–2071, Dec. 2013.

[19] B. Adams, D. Phung, and S. Venkatesh, "Extraction of social context and application to personal multimedia exploration," in *Proc. ACM Multimedia*, 2006, pp. 987–996.

[20] J. Zhuang, T. Mei, S. C. H. Hoi, X.-S. Hua, and S. Li, "Modeling social strength in social media community via kernel-based learning," in *Proc. ACM Multimedia*, 2011, pp. 113–122.

[21] T. Mei, X.-S. Hua, L. Yang, and S. Li, "VideoSense: Towards effective online video advertising," in *Proc. ACM Multimedia*, 2007, pp. 1075–1084.

[22] T. Mei, X.-S. Hua, and S. Li, "Contextual in-image advertising," in *Proc. ACM Multimedia*, 2008, pp. 439–448.

[23] T. Mei and X.-S. Hua, "Contextual Internet multimedia advertising," *Proc. IEEE*, vol. 98, no. 8, pp. 1416–1433, Aug. 2010.

[24] L. Li, T. Mei, and X.-S. Hua, "GameSense: Game-like in-image advertising," *Multimedia Tools Appl.*, vol. 49, no. 1, pp. 145–166, 2010.

[25] W.-S. Liao, K.-T. Chen, and W. H. Hsu, "AdImage: Video advertising by image matching and ad scheduling optimization," in *Proc. SIGIR*, 2008, pp. 767–768.

[26] T. Mei, J. Guo, X.-S. Hua, and F. Liu, "AdOn: Toward contextual overlay in-video advertising," *Multimedia Syst.*, vol. 16, nos. 4–5, pp. 335–344, 2010.

[27] T. Mei, X.-S. Hua, and S. Li, "VideoSense: A contextual in-video advertising system," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 12, pp. 1866–1879, Dec. 2009.

[28] T. Mei, L. Yang, X.-S. Hua, H. Wei, and S. Li, "VideoSense: A contextual video advertising system," in *Proc. ACM Multimedia*, 2007, pp. 463–464.

[29] C.-H. Chang, K.-Y. Hsieh, M.-C. Chung, and J.-L. Wu, "ViSA: Virtual spotlighted advertising," in *Proc. ACM Multimedia*, 2008, pp. 837–840.

[30] T. Li, T. Mei, I.-S. Kweon, and X.-S. Hua, "Multi-video synopsis for video representation," *Signal Process.*, vol. 89, no. 12, pp. 2354–2366, 2009.

[31] Y. Li, K. W. Wan, X. Yan, and C. Xu, "Real time advertisement insertion in baseball video based on advertisement effect," in *Proc. ACM Multimedia*, 2005, pp. 343–346.

[32] H. Liu, S. Jiang, Q. Huang, and C. Xu, "A generic virtual content insertion system based on visual attention analysis," in *Proc. ACM Multimedia*, 2008, pp. 379–388.

[33] B. N. Schenkman and F. U. Jönsson, "Aesthetics and preferences of Web pages," *Behaviour Inf. Technol.*, vol. 19, no. 5, pp. 367–377, 2000.

[34] N. Tractinsky, A. Cokhavi, M. Kirschenbaum, and T. Sharfi, "Evaluating the consistency of immediate aesthetic perceptions of Web pages," *Int. J. Human-Comput. Stud.*, vol. 64, no. 11, pp. 1071–1083, 2006.

[35] G. Buscher, S. T. Dumais, and E. Cutrell, "The good, the bad, and the random: An eye-tracking study of ad quality in Web search," in *Proc. ACM SIGIR*, 2010, pp. 42–49.

[36] R. Grier, P. Kortum, and J. Miller, "How users view Web pages: An exploration of cognitive and perceptual mechanisms," in *Human Computer Interaction Research in Web Design and Evaluation*. Hershey, PA, USA: IGI Global, 2007, pp. 22–41.

[37] Y. Yesilada, C. Jay, R. Stevens, and S. Harper, "Validating the use and role of visual elements of Web pages in navigation with an eye-tracking study," in *Proc. WWW*, 2008, pp. 11–20.

[38] R. A. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*. Reading, MA, USA: Addison-Wesley, 1999.

[39] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.

[40] T. Mei, L. Li, X.-S. Hua, and S. Li, "ImageSense: Towards contextual image advertising," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 8, no. 1, 2012, Art. no. 6.

[41] F. Nack, A. Manniesing, and L. Hardman, "Colour picking: The pecking order of form and function," in *Proc. ACM Multimedia*, 2003, pp. 279–282.

[42] O. Wu, Y. Chen, B. Li, and W. Hu, "Evaluating the visual quality of Web pages using a computational aesthetic approach," in *Proc. WSDM*, 2011, pp. 337–346.

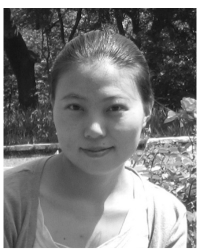
Tao Mei (M'07–SM'13) received the B.E. and Ph.D. degrees from the University of Science and Technology of China, Hefei, China, in 2001 and 2006, respectively.

He is currently a Lead Researcher with Microsoft Research, Beijing, China. He has authored or co-authored over 100 papers in journals and conferences and ten book chapters and has edited four books. He holds over 15 U.S. granted patents and over 20 pending. His current research interests include multimedia analysis and retrieval and computer vision.

Dr. Mei is a Senior Member of the ACM and a fellow of IAPR. He has received several paper awards from prestigious multimedia journals and conferences, including the IEEE Communications Society MMTC Best Journal Paper Award in 2015, the IEEE Circuits and Systems Society Circuits and Systems for Video Technology Best Paper Award in 2014, the IEEE TRANSACTIONS ON MULTIMEDIA Prize Paper Award in 2013, and Best Paper Awards at ACM Multimedia in 2009 and 2007. He was the Principle Designer of the automatic video search system that achieved the best performance in the worldwide TRECVID evaluation in 2007. He is an Editorial Board Member of the IEEE TRANSACTIONS ON MULTIMEDIA, *ACM Transactions on Multimedia Computing, Communications, and Applications*, *Machine Vision and Applications*, and *Multimedia Systems*. He was an Associate Editor of *Neurocomputing* and a Guest Editor of eight international journals. He is the General Co-Chair of ACM ICIMCS in 2013, the IEEE ICME 2015, the IEEE MMSP in 2015, and MMM in 2013. He will be the Program Co-Chair of ACM Multimedia in 2018 and has been the Area Chair for a dozen international conferences.

Lusong Li received the Ph.D. degree in computer software and theory from Beihang University, Beijing, China, in 2010.

He was a Research Fellow with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore. He is currently a Data Scientist and an Engineering Director of Graymatics Inc., Santa Clara, CA, USA. His current research interests include computer vision, machine learning, video content analysis, multimedia search, management, mining, advertising, and mobile multimedia computing.



Xinmei Tian (M'13) received the B.E. and Ph.D. degrees from the University of Science and Technology of China, Hefei, China, in 2005 and 2010, respectively.

She is currently an Associate Professor with the CAS Key Laboratory of Technology in Geo-spatial Information Processing and Application, University of Science and Technology of China. Her current research interests include multimedia information retrieval and machine learning.

Dr. Tian received the Excellent Doctoral Dissertation of the Chinese Academy of Sciences Award in 2012 and the Nomination of National Excellent Doctoral Dissertation Award in 2013.

Dacheng Tao (F'14) is currently a Professor of Computer Science with the Centre for Quantum Computation and Intelligent Systems and the Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney NSW, Australia. He mainly applies statistics and mathematics to data analytics. His current research interests include computer vision, data science, image processing, machine learning, neural networks, and video surveillance. His research results have resulted in one monograph and over 100 publications in prestigious journals and prominent conferences, such as the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the IEEE TRANSACTIONS ON IMAGE PROCESSING, the *Journal of Machine Learning Research*, the *International Journal of Computer Vision*, the Conference on Neural Information Processing Systems, the International Conference on Machine Learning, the Conference on Computer Vision and Pattern Recognition, the International Conference on Computer Vision, the European Conference on Computer Vision, the International Conference on Artificial Intelligence and Statistics, the International Conference on Data Mining (ICDM), and the ACM Special Interest Group on Knowledge Discovery and Data Mining.

Mr. Tao has received several best paper awards, such as the Best Theory/Algorithm Paper Runner Up Award at IEEE ICDM'07, the Best Student Paper Award at IEEE ICDM'13, and the 2014 ICDM 10 Year Highest-Impact Paper Award.

Chong-Wah Ngo received the B.Sc. and M.Sc. degrees in computer engineering from Nanyang Technological University, Singapore, and the Ph.D. degree in computer science from The Hong Kong University of Science and Technology, Hong Kong.

He was a Post-Doctoral Scholar with the Beckman Institute, University of Illinois at Urbana–Champaign, Champaign, IL, USA, and a Visiting Researcher with Microsoft Research Asia, Beijing, China. He is currently a Professor with the Department of Computer Science, City University of Hong Kong, Hong Kong. His current research interests include video computing and multimedia mining.

Dr. Ngo is an Associate Editor of the IEEE TRANSACTIONS ON MULTIMEDIA.