Neurocomputing 120 (2013) 524–535

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Position-wise contextual advertising: Placing relevant ads at appropriate positions of a web page

Zongda Wu [a,b], Guandong Xu [c], Chenglang Lu [a,d,*], Enhong Chen [b], Yanchun Zhang [e], Hong Zhang [f]

[a] Wenzhou University, Wenzhou 325035, Zhejiang, China
[b] University of Science and Technology of China, Hefei 230026, Anhui, China
[c] University of Technology, Sydney, Australia
[d] Northwestern Polytechnical University, Xi'an 710072, China
[e] Victoria University, Melbourne, Australia
[f] Institute of Science and Technology Information of Zhejiang Province, Hangzhou 310006, China

ARTICLE INFO

Article history:
Received 4 July 2012
Received in revised form
7 April 2013
Accepted 7 April 2013
Communicated by Tao Mei
Available online 15 June 2013

Keywords:
Wikipedia knowledge
Similarity
Contextual advertising

ABSTRACT

Web advertising, a form of online advertising, which uses the Internet as a medium to post product or service information and attract customers, has become one of the most important marketing channels. As one prevalent type of web advertising, contextual advertising refers to the placement of the most relevant ads at appropriate positions of a web page, so as to provide a better user experience and increase the user's ad-click rate. However, most existing contextual advertising techniques only take into account how to select as relevant ads for a given page as possible, without considering the positional effect of the ad placement on the page, resulting in an unsatisfactory performance in ad local context relevance. In this paper, we address the novel problem of position-wise contextual advertising, i.e., how to select and place relevant ads properly for a target web page. In our proposed approach, the relevant ads are selected based on not only global context relevance but also local context relevance, so that the embedded ads yield contextual relevance to both the whole target page and the insertion positions where the ads are placed. In addition, to improve the accuracy of global and local context relevance measure, the rich wikipedia knowledge is used to enhance the semantic feature representation of pages and ad candidates. Last, we evaluate our approach using a set of ads and pages downloaded from the Internet, and demonstrate the effectiveness of our approach.

1. Introduction

Web advertising is becoming an increasingly important and popular advertising market today. PwC[1] predicts that web advertising will become the second largest advertising medium in America after TV within the next four years, and spending in this area will increase from 24 billion dollars in 2009 to 34 billion dollars in 2014. A large part of web advertising consists of textual ads, which are short text messages usually marked as "sponsored links" or similar. Now, there are two main types of textual web advertising, i.e., sponsored search and contextual advertising [1,2]:

1. *Sponsored search* (also called keyword-targeted advertising), which selects ads based on keywords contained in search queries given by users, is characterized by placing paid textual ads links on the result pages returned by a web search engine (e.g., Google).
2. *Contextual advertising* (also called content-targeted advertising), which judges the context relevance of ads to the page that the user is browsing, refers to the selection of relevant commercial ads for the target page.

One of the important advantages of contextual advertising over the sponsored search is that it can support various types of web sites, which range from individual bloggers and small niche communities to large publishers (e.g., major newspapers). Now, almost all for-profit non-transactional sites, i.e., the sites that do not sell anything directly, rely heavily on the revenues from contextual advertising. Without contextual ads, the Web will lose the most of its market value.

The first major contextual advertising platform was provided by Google in 2003 [3]. Now, almost all popular search engines such as Baidu, Yahoo! and Microsoft Bing provide similar platforms for ad

* Corresponding author at: Wenzhou University, Wenzhou 325035, China.
E-mail addresses: zongda1983@163.com (Z. Wu), guandong.xu@vu.edu.au (G. Xu), chenglang.lu@qq.com (C. Lu), cheneh@ustc.edu.cn (E. Chen).

[1] PricewaterhouseCoopers, a global professional services company —http://www.pwc.com.

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.04.018

publishers and web site owners. As shown in Fig. 1, a contextual advertising platform generally consists of the following four parts [1,2,4–6]:

1. The *advertiser* provides the supply of textual ads, which is usually a company that wants to use the ad platform to promote their products or services, and needs to pay for its textual ads.
2. The *publisher* is the owner of a web site on which textual ads are placed, who typically aims to provide a good user experience and increase the number of ad-clicks, so as to maximize the market revenue.
3. The *ad platform* is a software system of matching ads to pages, which allows the publisher to select appropriate ads based on the context similarity between pages and ads for advertisers that create the ad revenue for the publisher.
4. End *users* consist of customer groups who have potential interest in the ads while browsing the content of a web page, supplied by the publisher.

The most dominant online advertising pricing model is pay-per-click (PPC), where the advertisers pay a certain amount to the publisher and the ad platform for each user's click on the ads. In addition, there are also other types of pricing models for textual ads, including: (1) pay-per-impression (PPI), where the advertisers pay for the number of ads displayed on a web page and (2) pay-per-action (PPA), where the payment made by the advertiser is calculated by each sale originating from the ads. Since most existing contextual advertising approaches are based on

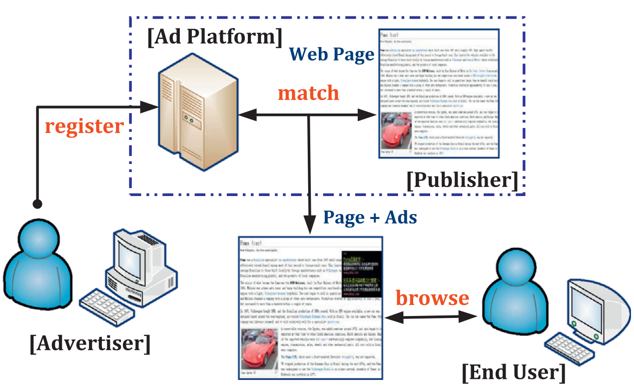


Fig. 1. A contextual advertising platform.

the PPC model [1,2,4–6], in this paper, we also use this model for simplicity.

1.1. Problem statement

Under the PPC pricing model, in [2,6,7], it has been pointed out that given a target page $\boldsymbol{p}$, the revenue of the publisher and the ad platform can be estimated as: $\sum_{i=1,2,\ldots,k} P(click|\boldsymbol{p}, a_i)\cdot price(a_i)$, where k is the number of ads displayed on the page $\boldsymbol{p}$ and $price(a_i)$ is the click-price of the current ad a_i. After simplifying the model, the revenue can be maximized by searching for: $\arg\max_i P(click|\boldsymbol{p}, a_i)$. Thus, to maximize the PPC pricing model, we would need to select proper ads to maximize the probability of user's ad-clicks.

In general, the probability of user's ad-clicks is positively related to user's interest in the ad content. Furthermore, we assume that a web user should be interested in the page content that he is currently browsing, i.e., the page content that a user currently browsing could reflect the user's interest to a certain extent. Thus, we believe that the users' ad-click rate can be boosted by increasing the context relevance of an ad to the page position where the ad is placed. A number of studies also have shown similar conclusions, e.g., it was pointed out in [2,6,7] that the ads should be relevant to their surrounding page content; and it was further pointed out in [8–10] that the in-image ads should be not only relevant to the entire page but also locally relevant to their hosting image.

Based on the above observations, to maximize the PPC pricing model (i.e., to increase the users' ad-click rate as much as possible, and thus maximize the revenue), we would need to solve two key problems. The *first problem* is how to select as relevant ads for a given page as possible. As shown in Fig. 2(a) and (b), for example, given a page about "travel in China", embedding the page with an ad about "hotel information" or "tour service" would attract more user attention than randomly chosen ads.

A web page (especially for a long page), however, may cover multiple topics (or multiple subtopics under the same topic), such that an ad, which is relevant to the entire page, may be not locally relevant to the page position where the ad is placed. The *second problem* is the placement of relevant ads at appropriate positions of a web page, making each ad close to the page segments that are really relevant to the ad. For example, as shown in Fig. 2(b) and (c), for the page position next to the segment "Beijing", embedding the position with the ad "Great Wall tour" would increase the interest of users clicking on the ad more than the other ad "Xian hotel", although the two ads are both relevant to the entire page.

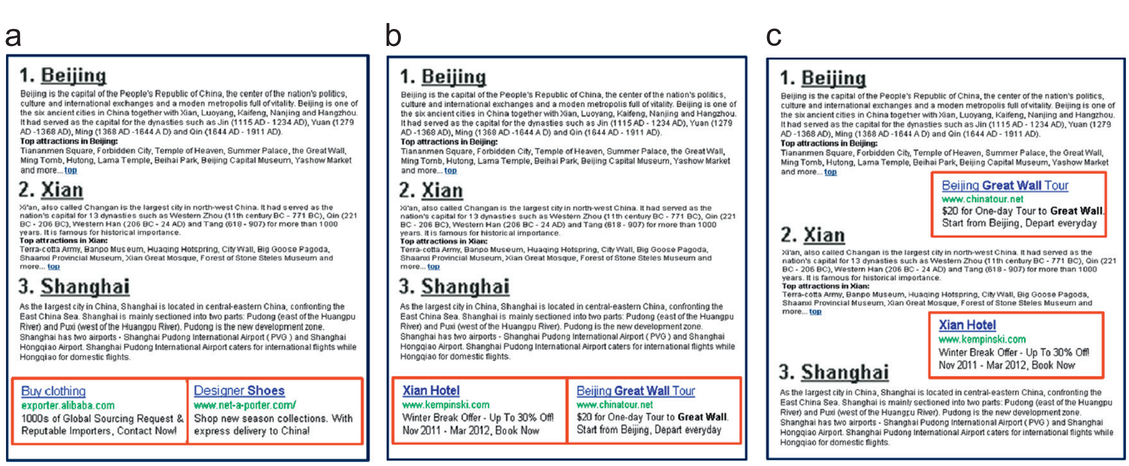


Fig. 2. Three examples of placing ads on a web page, where the highlighted areas indicate associated ads.

Now, there have been a number of studies addressing contextual advertising (see the related work given in Section 2), however, most of which estimated the context relevance by analyzing the co-occurrence of the same keywords or phrases between pages and ads (namely, based on keyword matching). On the one hand, as pointed out in [4], the keyword matching may cause problems such as homonymy and polysemy, low intersection of keywords and context mismatch, resulting in the selection of irrelevant ads for a target web page and the decrease of accuracy on contextual advertising. On the other hand, all these studies mainly focus on how to select globally relevant ads for a given web page, without considering how to place the ads on the web page properly (i.e., the selected ads would be always embedded into a predefined position), making the embedded ads locally irrelevant to the surrounding page content sometimes, consequently, decreasing the willingness of users to click the ads to a certain extent. To the best of our knowledge, our paper is one of the first attempts to explore the placement of relevant ads in a web page (i.e., position-wise contextual advertising).

1.2. Contributions

Motivated by the above observations, in this paper we propose a novel contextual advertising approach by considering not only the selection of relevant ads for a target web page but also the placement of the ads on the page. More specifically, the contributions of this paper include the following three aspects:

1. A *framework* of considering both global and local context relevance is proposed, so that the selected ads yield context relevance to both the whole target page and the insertion positions where the ads are placed.
2. A *context relevance measure approach* is proposed by leveraging wikipedia, one of the largest human knowledge bases, to enrich the semantic feature representation for pages and textual ads, so as to improve the measure accuracy of global and local context relevance.
3. Last, we perform *experimental evaluation* in terms of advertising accuracy and running efficiency over a set of web pages and ad candidates, which are downloaded from the Internet.

1.3. Organizations

The rest of this paper is organized as follows. In Section 2, we survey the related work about contextual advertising techniques. Section 3 formulates the problem of position-wise contextual advertising. Section 4 describes our proposed approach to position-wise contextual advertising, i.e., based on the wikipedia knowledge, how to select and place relevant ads properly for a target web page. Section 5 presents our experimental setup and experimental evaluation results. Section 6 discusses the novelty of our proposed approach as well as some disadvantages. Last, we conclude our work in Section 7.

2. Related work

In this section, we review the literatures related to our work from the following two aspects: keyword matching and wikipedia matching.

2.1. Keyword matching

Traditional keyword matching can be used to estimate the ad relevance by analyzing the co-occurrence of the same keywords between the ad and the page. In this approach, the cosine measure [11] is used to calculate the similarity between a page and an ad. The cosine measure is an important technique often used in information retrieval to calculate the similarity between textual documents. In addition, the technique about how to implement an effective content extraction from a web page (mainly referring to a HTML document) is also important, which has been discussed in [12].

A recent research result about applying keyword matching into contextual advertising was presented in [13]. In this work, all the ads and pages are represented as vectors in a vector space. In order to solve the problem of the low intersection of keywords between pages and ads, the authors proposed to augment a page with additional keywords from other pages that are similar to that page. Next, the authors explored 10 different strategies to select different parts of pages and ads, used as a basis for the vectors of the pages and ads. Last, the authors matched the pages and the ads based on the cosine of the angle between the ad vector and the page vector to select relevant ads for a given target page.

In [14], under the assumption that an ad can be viewed as a noisy translation of a web page, the authors selected the ads that provide the best translation for the page. In [6], an approach to leveraging sentiment detection to improve contextual advertising was presented, which combines contextual advertising matching with sentiment analysis to select ads that are related to the positive (or neutral) aspects of a blog and ranks the ads according to their relevance. In [15], the authors proposed to use lexical graphs created from web corpora as a means of computing improved content similarity metrics between ads and pages. The results indicated that using lexical graph can provide evidence of significant improvement in the perceived relevance of the recommended ads. A new architecture (called blog context overlay network) was proposed in [16], to fulfill context matching between blog-based knowledge management systems, and as a conclusion, a measurement for contextual similarity between blogs was also presented.

In [8,9], aiming at contextual image advertising, the authors proposed an advertising system driven by images, which can automatically associate relevant ads with an image rather than the entire text on a page, and seamlessly insert the ads in non-intrusive areas within each individual image. In this system, the relevant ads were selected based not only on textual relevance but also on visual similarity, so that the ads yield contextual relevance to both the text in the page and the image content. Moreover, the ad insertion positions were detected based on image saliency to minimize intrusiveness. The authors also used traditional keyword matching for ad selection.

As pointed out in [4,17], however, the main drawback of keyword matching is that it may lead to some problems such as homonymy and polysemy, low intersection of keywords and context mismatch, resulting in dramatically degrading the context relevance of selected ads to their pages.

2.2. Wikipedia matching

In order to solve the problems of homonymy and polysemy, the low intersection of keywords and context mismatch, in the area of clustering, a new technique called wikipedia matching was proposed recently, which uses wikipedia as the reference model to enhance the semantic representation of text documents and thus improve the accuracy of similarity measure of text documents. The basic idea of wikipedia matching is to introduce wikipedia articles as a semantic reference space, and in turn, the semantics of each text document are reflected and enhanced by projecting the original term space of text documents into this additional reference model.

In [4], based on wikipedia matching, the authors presented a solution to contextual advertising. In their work, 1000 feature articles are first chosen from wikipedia. Next, for each ad, the feature articles that are related to the ad are selected by using the traditional cosine similarity measure; and for a page, the same procedure is followed and articles which are related to the page are selected. Last, using the feature articles as the intermediate semantic reference model on which the ads and the page are re-expressed as vectors, the approach determines the ads that exhibit more relevance to the page, and construct a ranking function to select the ads that are most relevant to the page. In [18,19], the authors improved this work by introducing three selective matching strategies to refine the selection of really relevant reference articles for a page (or an ad), so as to better balance accuracy and efficiency in contextual advertising.

In [17,20], the authors proposed a similar approach, aiming at textual document clustering. First, the authors automatically construct a thesaurus of concepts from wikipedia. Then, they introduce a framework to expand the traditional representation of document terms with semantic relations (i.e., synonymy, hyponymy and associative relations), demonstrating its efficacy in enhancing previous methods for text classification.

In [21], the authors proposed to use wikipedia, one of the largest human knowledge base, to understand a user's query intent, without the need to collect large quantities of examples to train an intent classifier. In this approach, the wikipedia concepts are used as the intent representation space, thereby, each intent domain is represented as a set of wikipedia articles and categories. Then, the intent of any input query is identified through mapping the query into the wikipedia representation space. Compared to previous work, this approach can achieve better coverage to classify queries in an intent domain, although the number of seed intent examples is small.

However, most existing wikipedia matching techniques are not designed for contextual online advertising, resulting in seldom considering the problem of running efficiency. Therefore, these techniques are generally difficult to be applied into contextual advertising directly. In addition, the main drawback of most existing contextual advertising techniques (including keyword matching and wikipedia matching), except what mentioned above, is that they only focus on the selection of relevant ads for a target web page, without considering the positional effect of the ad placement on the page.

3. Problem definition

Without loss of generality, the problem of position-wise contextual advertising is defined as the selection and placement of the most relevant ads for a given target web page. In this section, we provide a formal definition of this problem.

Let $\boldsymbol{p}$ be a target page that consists of N_b ad insertion positions, represented by $\mathcal{B}=\{b_i\}_{i=1}^{N_b}$. Let $\mathcal{A}$ be the candidate ad database that contains N_a ads, represented by $\mathcal{A}=\{a_j\}_{j=1}^{N_a}$. Let N be the number of expected ads to be embedded into a page, generally, which is given by the publisher. Let $R_G(\boldsymbol{p}, a_j)$ be the global context relevance between the page $\boldsymbol{p}$ and the ad a_j, and $R_L(b_i, a_j)$ be the local context relevance between the ad insertion position b_i and the ad a_j. Suppose we introduce the following design variables: $\boldsymbol{x}=[x_1, x_2, \ldots, x_{N_a}]$ and $\boldsymbol{y}=[y_1, y_2, \ldots, y_{N_b}]$, where x_j and y_i indicate whether a_j and b_i are selected ($x_j=1$ and $y_i=1$) or not ($x_j=0$ and $y_i=0$). Then, the above expectation about selecting the most relevant N ads from the database $\mathcal{A}$ and placing them properly on the page $\boldsymbol{p}$ can be formulated as the following nonlinear 0–1 integer programming problem (NIP) [22]:

$$\max_{(\boldsymbol{x},\boldsymbol{y})} f(\boldsymbol{x},\boldsymbol{y}) = w_g \sum_{j=1}^{N_a} x_j R_G(\boldsymbol{p}, a_j) + w_l \sum_{j=1}^{N_a} \sum_{i=1}^{N_b} x_j y_i R_L(b_i, a_j)$$

$$\text{s.t.} \quad \sum_{j=1}^{N_a} x_j = N, \quad \sum_{i=1}^{N_b} y_i = N, \quad x_j, y_i \in \{0,1\} \tag{1}$$

In the above equation, the parameters w_g and w_l control the emphasis on global and local context relevance, and satisfy the constraints: $1 \geq w_g, w_l \geq 0$ and $w_g + w_l = 1$, whose suitable values can be trained by cross-validation experiments.

Based on Eq. (1), we can conclude that the effectiveness of a position-wise contextual advertising approach mainly depends on the global and local context relevance $R_G(\boldsymbol{p}, a_j)$ and $R_L(b_i, a_j)$, whose accuracy and efficiency would directly determine the accuracy and efficiency of position-wise contextual advertising.

4. Methodology

In this section, we present our approach to position-wise contextual advertising. In this work, we take into account comprehensively both global and local context relevance, so as to improve the overall ad's relevance to the target web page; and we combine informative wikipedia knowledge with conventional keyword matching, so as to improve the measure accuracy of context relevance. The main framework of our approach is presented in Fig. 3, which comprises the following four steps:

1. *Page segmentation*: for a target page, a page segmentation algorithm is used to prune all unnecessary parts (such as navigation links and HTML elements) and partition the page into several segments.
2. *Page featured representation*: based on the informative wikipedia knowledge, each segment is mapped into three feature representation vectors, a keyword vector, a concept vector and a category vector.
3. *Ad insertion position detection*: all ad insertion positions of the page are detected based on the context relevance among all the

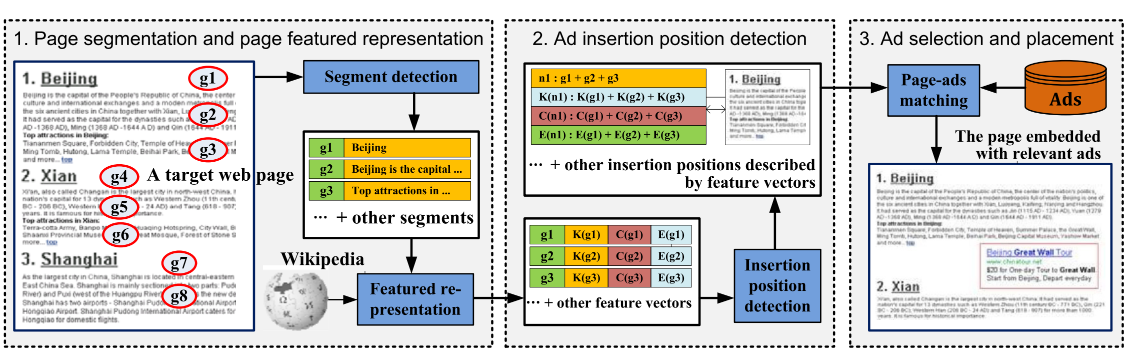


Fig. 3. The main framework of our approach to position-wise contextual advertising.

segments, such that in each insertion position a segment should be of higher internal relevance with other segments in the same block, but of lower external relevance with those in different blocks. Meanwhile, for each ad insertion position, its feature vectors are generated by combining the feature vectors of all the segments in the block.

4. *Ad selection and placement*: the page-ads matching module ranks all candidate ads according to the global and local relevance between the ads and the content of the whole page and the insertion positions, such that the ads that yield context relevance to both the page and the insertion positions would be embedded into the target page.

It should be pointed out that each ad in the ad candidate database has been in advance mapped into feature vectors (i.e., a keyword vector, a concept vector and category vector). Besides, we also make the preprocessing of wikipedia concept and category information to be completed offline. Thus, the total actual running cost of our approach, which is mainly dependent on the time spent generating the required feature vectors for the target page and making the context matching between the page and all ad candidates, would be dramatically decreased to improve the real advertising performance. In the following sections, we will detail these processing steps, respectively.

4.1. Page segmentation

This step aims to complete the preparation work for subsequent ad selection and placement. First, each target page $\boldsymbol{p}$ is processed by an HTML content extraction tool [23], which can prune all unnecessary parts for a web page (such as navigation links and HTML elements) by analyzing an HTML DOM tree, consequently, keeping only the main text content (instead of the whole page). Second, based on paragraph partition flags (such as newline), the main content text of the page $\boldsymbol{p}$ is partitioned into several segments, represented by $\mathcal{G}=\{g_i\}_{i=1}^{N_g}$. Third, non-keywords without concrete meaning (e.g., articles, prepositions) are removed from each segment so as to improve the context relevance measure accuracy. This process is completed using a reference library of non-keywords. In doing this, we download a word library from Youdao Dictionary[2] that consists of about 82 730 common words, and find out all the function words from the dictionary as the reference library of non-keywords. Last, each keyword of segments is processed by a stemming algorithm [23] that truncates the suffix of the word and reduces it to a stem. It should be pointed out that since this step adopts a simple page content extraction tool, it cannot work well with regard to complex web pages, e.g., HTML + CSS + JavaScript (i.e., actually, we simply treat each page as a text document).

Now, we obtain a set of keywords for each segment g_i in the target page $\boldsymbol{p}$, represented by $\mathbf{K}(g_i)=\{\mathbf{count}(w_j,g_i)\}_{j=1}^{N_k^i}$, where $\mathbf{count}(w_j,g_i)$ denotes the frequency of occurrences of the keyword w_j in the segment g_i and $N_k{}^i$ denotes the number of keywords in the segment g_i. Next, based on $\mathbf{K}(g_i)$, combined with the *tf-idf* (*term frequency-inverse document frequency*) weight [24], we can further generate a feature representation vector (called a keyword vector) for the segment g_i, represented by $\mathbf{V_K}(g_i)$, which consists of the *tf-idf* values of all the keywords contained in the segment.

4.2. Page featured representation

After page segmentation, all segments are represented as “sets of keywords” (i.e., keyword vectors), and the textual closeness between two segments can be measured by the similarity of the two keyword vectors (i.e., keyword matching). However, as mentioned in the related work section, keyword matching, due to not taking into account any semantic information, may lead to some problems such as homonymy and polysemy, low intersection of keywords, context mismatch, etc., and thus impact the context relevance accuracy measured by using keyword matching alone.

In order to improve the measure accuracy of context relevance, we leverage the informative wikipedia knowledge along with keyword matching, to judge the overall semantic and textual similarity between segments (as well as pages and ad candidates). To do this, this step aims to map each page segment into two new feature representation vectors: a concept vector and a category vector, which would be used to measure the semantic similarity between segments. In addition, compared to existing relevance measure approaches, which are not designed for contextual advertising, resulting in seldom considering the problem of running efficiency, this step also aims to take into consideration both advertising accuracy and running efficiency, so as to make our approach more applicable for contextual advertising. Below, we describe the mapping from a segment to a concept vector and to a category vector, respectively.

4.2.1. Concept mapping

In wikipedia, each article describes one concept, and the concept is named after the title of the article, so if the title of a concept occurs in a segment, it is very likely that the concept is semantically related to the segment. First, we aim to discover all the concepts mentioned in a segment and count the frequency of occurrences of the concepts in the segment. To do this, we scan each segment $g_i \in \mathcal{G}$ to find out all the wikipedia titles that are mentioned in the segment g_i, consequently, obtaining a set $\mathcal{T}(g_i)$ of titles and a set $\mathcal{C}(g_i)$ of corresponding concepts (i.e., named after the titles). However, due to synonymy and polysemy, in wikipedia, a concept may correspond to several titles, and vice versa, i.e., it is a many–many mapping from $\mathcal{T}(g_i)$ to $\mathcal{C}(g_i)$. In order to solve the problem caused by ambiguous keywords (i.e., polysemy), we define the frequency of occurrences of each concept $c_j \in \mathcal{C}(g_i)$ related to any title $\boldsymbol{t} \in \mathcal{T}(g_i)$ in the segment g_i as follows:

$$\mathbf{count}(c_j,g_i)[\boldsymbol{t}] = \begin{cases} 0, & \text{if } \boldsymbol{t} \text{ is not a title of the wikipedia concept } c_j \\ \left(\dfrac{\mathbf{sim^k}(c_j,g_i)}{\sum_{u\in\mathcal{C}(\boldsymbol{t})}\mathbf{sim^k}(u,g_i)} \cdot \mathbf{count}(\boldsymbol{t},g_i), \quad \text{otherwise}\right) \end{cases} \tag{2}$$

where (1) $\mathbf{count}(\boldsymbol{t},g_i)$ denotes the frequency of occurrences of $\boldsymbol{t}$ in g_i; (2) $\mathcal{C}(\boldsymbol{t})$ denotes a set of concepts named after $\boldsymbol{t}$; and (3) $\mathbf{sim^k}(c_j,g_i)$ denotes the textual similarity between c_j and g_i, which is calculated by using traditional keyword matching. Then, the frequency of occurrences of the concept c_j in the segment g_i can be computed through summing up the frequency values of occurrences of the concept c_j related to each of its titles in $\mathcal{T}(g_i)$, i.e., $\mathbf{count}(c_j,g_i) = (\sum_{u\in\mathcal{T}(g_i)}\mathbf{count}(c_j,g_i)[u])$.

Now, we obtain a set of concepts for each segment g_i in the target page, represented by $\mathbf{C}(g_i)=\{\mathbf{count}(c_j,g_i)\}_{j=1}^{N_c^i}$ (where $N_c{}^i$ denotes the number of concepts mentioned in g_i), which consists of frequency of occurrences of each concept $c_j \in \mathcal{C}(g_i)$ in the segment g_i.

In the above process, synonymy keywords or phrases would be mapped into the same concept, thereby solving the context mismatch caused by synonyms. Moreover, for an ambiguous keyword, which may be associated with some different concepts, a fulltext comparison (see Eq. (2)) would be made between each of these concepts and the segment, to determine the concept

[2] Youdao Dictionary, a dictionary released by the NetEase company—http://dict.youdao.com

distribution represented in the segment in terms of various weights, thereby easing the problem of homonymy and polysemy.

Second, we aim to further expand $\mathbf{C}(g_i)$ by adding more semantically related concepts. As described in [21], given that there is a connection between two concepts (i.e., the interlink between each other) in wikipedia, it is likely that the two concepts share common topics, i.e., which are semantically related to each other. The expansion of $\mathbf{C}(g_i)$ is based on such an observation. Specifically, the expansion is a breadth-first graph traversal process that uses the concepts associated with $\mathbf{C}(g_i)$ as start traversal nodes: (1) for each concept c_j associated with $\mathbf{C}(g_i)$, we obtain all the concepts that connect to c_j in wikipedia, represented by $\mathcal{C}(c_j)$; (2) we calculate the frequency of occurrences of each concept $\boldsymbol{e} \in \mathcal{C}(c_j)$ in the segment g_i as follows:

$$\mathbf{count}\left(\boldsymbol{e}, g_i\right) = \left(\frac{\mathbf{num^k}(\boldsymbol{e}, c_j)}{\sum_{u \in \mathcal{C}(c_j)} \mathbf{num^k}(u, c_j)}\right) \cdot \mathbf{count}\left(c_j, g_i\right) \tag{3}$$

where $\mathbf{num^k}(\boldsymbol{e}, c_j)$ denotes the number of hyperlinks between the two concepts $\boldsymbol{e}$ and c_j; (3) if $\mathbf{count}(\boldsymbol{e}, g_i)$ is greater than a given threshold $\boldsymbol{\lambda_c}$ that is used to control the depth of graph traversal, then $\mathbf{count}(\boldsymbol{e}, g_i)$ will be added into $\mathbf{C}(g_i)$ and (4) the steps 1–3 are kept on until all the concepts (including those new added) in $\mathbf{C}(g_i)$ are traversed.

By using the above process, the feature representation of a page segment would be further enriched with more semantic concepts, each of which is not mentioned in the segment but semantically related to the segment to a certain degree (indirectly related), thereby, raising the intersection of concepts between segments, and easing the problem of the low intersection of keywords.

4.2.2. Category mapping

In wikipedia, category information provides additional thesaurus knowledge to reflect the semantic relationship between concepts. For a category in wikipedia, if there are some concepts that belong to the category and are semantically related to a segment, then it is very likely that the category is also semantically related to the segment. Based on such an observation, this step aims to leverage wikipedia category information to further enrich the semantic feature representation of segments.

Let $\mathcal{C}(e_j)$ denote a set of concepts belonging to a category e_j, and $\mathcal{E}(e_j)$ denote a set of all the immediate subcategories belonging to e_j (i.e., there is a hyperlink from each category in $\mathcal{E}(e_j)$ to e_j). $\mathcal{C}(e_j)$ and $\mathcal{E}(e_j)$ can be both determined by the hierarchical categorization structure in wikipedia. Then, the frequency of occurrences of the category e_j in a segment $g_i \in \mathcal{G}$ is defined as follows:

$$\mathbf{count}\left(e_j, g_i\right) = \left(\sum_{c \in \mathcal{C}(e_j)} \frac{\mathbf{count}(c, g_i)}{\theta_c}\right) + \left(\sum_{e \in \mathcal{E}(e_j)} \frac{\mathbf{count}(e, g_i)}{\theta_e}\right) \tag{4}$$

where $\boldsymbol{\theta}_c$ and $\boldsymbol{\theta}_e$ are two attenuation coefficients that are used to balance the importance of frequency values of categories in different depths. In our following experiments, empirically, we set $\boldsymbol{\theta_c} = 4.0$ and $\boldsymbol{\theta_e} = 8.0$.

Based on Eq. (4), a set $\mathbf{E}(g_i) = \{\mathbf{count}(e_j, g_i)\}_{j=1}^{N_e^i}$ of categories for the segment g_i is obtained, which consists of the frequency of occurrences of each category e_j in the segment g_i. The generation of $\mathbf{E}(g_i)$ is also a breadth-first graph traversal process that uses the concepts belonging to $\mathbf{C}(g_i)$ as start traversal nodes, where the traversed graph consists of (1) *nodes*, concepts and categories and (2) *edges*, the hyperlinks within concepts and categories, as well as the hyperlinks within categories. A threshold $\boldsymbol{\lambda_e}$ is also given to control the depth of such a graph traversal.

After the above process, the semantic feature representation for a segment would be further enriched with categories, each of which is semantically related to the segment more or less, thereby further easing the problems of homonymy and polysemy, low intersection of keywords, context mismatch, etc.

Now, for each segment g_i in the target page $\boldsymbol{p}$, we obtain a set $\mathbf{C}(g_i)$ of concepts and a set $\mathbf{E}(g_i)$ of categories. Based on $\mathbf{C}(g_i)$ and $\mathbf{E}(g_i)$, combined with the *tf-idf* weight, we can further generate two new feature representation vectors for the segment g_i: a keyword vector and a category vector, respectively, represented by $\mathbf{V_C}(g_i)$ and $\mathbf{V_E}(g_i)$. Besides, in the above process of concept mapping and category mapping, $\boldsymbol{\lambda_c}$ and λ_e are two important thresholds used to control the graph traversal depth: (1) a smaller value will achieve a better result on expanding concepts or categories but results in a longer search time; whereas (2) a larger value will lead to a shorter expansion result but a smaller running cost. In our following experiments, we use greater values (we set $\boldsymbol{\lambda_c} = 1$ and $\lambda_e = 0.5$) for the expansion on segments so as to obtain good running performance. This is due to the online featured representation for the target page, resulting in our prior consideration of efficiency.

4.3. Ad insertion position detection

Now, each segment in the target page has been represented as three feature vectors: a keyword vector, a concept vector and a category vector. This step aims to calculate the similarity between feature representation vectors to measure the context relevance between segments, and then merge adjacent segments with higher relevance into the same block as a candidate ad insertion position. In this process, we particularly take two aspects of similarity measures between segments into consideration: (1) the similarity between keyword vectors capturing the textual closeness and (2) the similarity between concept vectors (and category vectors) measuring the relevance from the semantic perspectives of thesaurus ontology.

For two segments g_i and g_j ($g_i, g_j \in \mathcal{G}$) in the target page $\boldsymbol{p}$, the overall textual and semantic similarity between them is defined as follows:

$$\begin{aligned}\mathbf{sim}(g_i, g_j) &= \underbrace{m_k \cdot \mathbf{sim}^k(g_i, g_j)}_{\textbf{textual similarity}} + \underbrace{m_c \cdot \mathbf{sim}^c(g_i, g_j) + m_e \cdot \mathbf{sim}^e(g_i, g_j)}_{\textbf{semantic similarity}} \\ &= m_k \cdot \cos \angle \mathbf{V_K}(g_i), \mathbf{V_K}(g_j) \\ &\quad + m_c \cdot \cos \angle \mathbf{V_C}(g_i), \mathbf{V_C}(g_j) + m_e \cdot \cos \angle \mathbf{V_E}(g_i), \mathbf{V_E}(g_j)\end{aligned} \tag{5}$$

where the parameters (m_k, m_c, m_e) are used to control the weight of concept vectors and category vectors in the semantic similarity computation between segments, as well as the balance between textual similarity and semantic similarity, which satisfy the constraints: $1 \geq (m_k, m_c, m_e) \geq 0$ and $(m_k + m_c + m_e) = 1.0$. It can be seen from the above equation that the similarity measure between segments is based on the cosine metric [12], namely the similarity between two segments is measured by the cosine value of the angle between the feature vectors of the segments.

Now, by using Eq. (5), the context relevance between any two segments in the target page can be measured accurately. Then, we cluster iteratively adjacent segments in the target page into the same block if their context relevance is greater than a given threshold, consequently making that all the segments clustered in the same block are of higher relevance to each other but of lower relevance to others in different blocks. As a result, we obtain a set of blocks $\mathcal{B}=\{b_i\}_{i=1}^{N_b}$, used as candidate ad insertion positions. Algorithm 1 details the above process.

Now, we complete the feature representation for the target page $\boldsymbol{p}$, each ad $a_j \in \mathcal{A}$ and each ad insertion position $b_i \in \mathcal{B}$. Then, the global context relevance between a_j and $\boldsymbol{p}$, and the local context relevance between a_j and b_i can be given by

$$R_G(\boldsymbol{p}, a_j) = \mathbf{sim}(\boldsymbol{p}, a_j) = m_k \cdot \cos \angle \mathbf{V_K}(\boldsymbol{p}), \mathbf{V_K}(a_j) + m_c \cdot \cos \angle \mathbf{V_C}(\boldsymbol{p}), \mathbf{V_C}(a_j) + m_e \cdot \cos \angle \mathbf{V_E}(\boldsymbol{p}), \mathbf{V_E}(a_j) \quad (6)$$

$$R_L(b_i, a_j) = \mathbf{sim}(b_i, a_j) = m_k \cdot \cos \angle \mathbf{V_K}(b_i), \mathbf{V_K}(a_j) + m_c \cdot \cos \angle \mathbf{V_C}(b_i), \mathbf{V_C}(a_j) + m_e \cdot \cos \angle \mathbf{V_E}(b_i), \mathbf{V_E}(a_j) \quad (7)$$

From Eqs. (6) and (7), it can be seen that, similar to Eq. (5), the global and local context relevance between an ad and a page are measured by the overall textual and semantic similarity, so as to improve the measure accuracy.

After substituting Eqs. (6) and (7) into Eq. (1), we can solve the problem of position-wise contextual advertising, which is formally defined in Section 3. However, the number of solutions in total to Eq. (1) is equal to $(C_{N_a}^N \cdot C_{N_b}^N \cdot N!)$, and the solution searching space will increase dramatically with the increase of the number of elements in $\mathcal{A}$ and $\mathcal{B}$. For example, if there are 10 000 ads in $\mathcal{A}$, 10 placing positions in $\mathcal{B}$, and three ads embedded

Algorithm 1. Ad insertion position detection.

Input: (1) $\mathcal{G}$, a set of segments contained in a target page, (2) each segment $g_i \in \mathcal{G}$ followed with three feature representation vectors, i.e., $\mathbf{V_K}(g_i)$, $\mathbf{V_C}(g_i)$ and $\mathbf{V_E}(g_i)$, and (3) a threshold $\boldsymbol{\lambda_t}$, which is set to 0.4 in our experiments.
Output: $\mathcal{B}$, a set of ad insertion positions, each comprised of related segments.

```
begin
  set B = empty;  set flag = true; // note: initialization.
  while flag = true do
    for i from 1 to ||G|| − 1 do
      if sim(g_i, g_{i+1}) ≥ λ_t then
        merging the two segments g_i and g_{i+1} into the same block, noted as g;
        recomputing the feature vectors for the block g_t by combining the feature
        vectors of the segments g_i and g_{i+1} together, consequently obtaining V_K(g),
        V_C(g) and V_E(g);
        adding the new generated block g into B.
      else
        adding the two segments g_i and g_{i+1} into B directly.
    if G = B then
      set flag = false. // note: if any two segments have not been merged together, set
      flag to terminate the iteration.
    else
      set G = B;  set B = empty. // note: otherwise, reset G and keep on the iteration.
  return B. // note: each element b_i ∈ B followed with V_K(b_i), V_C(b_i) and V_E(b_i).
```

4.4. Ad selection and placement

After ad insertion position detection, the target web page $\boldsymbol{p}$ has been divided into some relatively independent "insertion points". Based on the feature vectors of the ad insertion positions, we can further generate feature vectors for the target page $\boldsymbol{p}$, represented by $\mathbf{V_K}(\boldsymbol{p})$, $\mathbf{V_C}(\boldsymbol{p})$ and $\mathbf{V_E}(\boldsymbol{p})$.

Moreover, for each candidate ad $a_j \in \mathcal{A}$, we have in advance generated its feature vectors, noted as $\mathbf{V_K}(a_j)$, $\mathbf{V_C}(a_j)$, and $\mathbf{V_E}(a_j)$, by following the similar procedure to segment feature representation. However, this step is completed offline, without the need to consider the running performance. Thus, at the process of concept mapping and category mapping, different to segment feature representation, we set smaller values for the thresholds $\boldsymbol{\lambda_c}$ and $\boldsymbol{\lambda_e}$ (which are set to 0.02 and 0.05, respectively), so as to expand ads of relatively limited size with more semantically related concepts and categories, and as a result ease the problem of low intersection of keywords.

in a page, then the number of solutions will be nearly 1.2×10^{14}, which is extremely large. To solve this problem, we use the genetic algorithm (GA) to search solutions approaching the global optimum. Practically, we use an algorithm similar to that in [9], which is shown as Algorithm 2. As a result, the number of possible solutions is reduced to $(N_a \log N_a + N_b N'_a + N_b \log N_b)$.

Algorithm 2. Ad selection and placement.

Input: (1) $\boldsymbol{p}$, a target page and its feature vectors, (2) $\mathcal{B}$, a set of ad insertion positions, each element followed with feature vectors, and (3) $\mathcal{A}$, a set of candidate ads, each element followed with feature vectors.
Output: $\mathcal{R}'$, a solution approaching the global optimum to the problem of position-wise contextual advertising (i.e. Equation (1)).

begin
- **set** $\mathcal{R}$ = empty; // **note**: initialization.
- rank all the ad candidates in $\mathcal{A}$ according to the global context relevance $R_G(\boldsymbol{p}, a_j)$ between the target page $\boldsymbol{p}$ and each ad $a_j \in \mathcal{A}$ in a descendent order, and then select the top N'_a ads ($N < N'_a \ll N_a$), represented by $\mathcal{A}'$;
- **for** each ad insertion position $b_i \in \mathcal{B}$ **do**
  - from $\mathcal{A}'$, select the ad candidate $\boldsymbol{a}$ with the largest local context relevance to the position b_i, i.e., $\forall a_j \, (a_j \in \mathcal{A}' \to R_L(b_i, \boldsymbol{a}) \geq R_L(b_i, a_j))$;
  - remove the ad $\boldsymbol{a}$ from $\mathcal{A}'$;
  - add the pair $(b_i, \boldsymbol{a})$ into $\mathcal{R}$.
- rank each element $(b_i, a_j) \in \mathcal{R}$ according to the global context relevance $R_G(\boldsymbol{p}, a_j)$ plus the local context relevance $R_L(b_i, a_j)$ in a descendent order, and then select the top N elements, represented by $\mathcal{R}'$;
- **return** $\mathcal{R}'$. // **note**: each pair $(b_i, a_j) \in \mathcal{R}'$ denotes that the ad a_j would be selected and placed at the position b_i of the page $\boldsymbol{p}$.

5. Experiments

In this section, we experimentally evaluate our approach. First of all, we present data and the evaluation methodology, and several candidates which are used as a comparison with our approach. Then, we evaluate our approach from the following three aspects: (1) time spent selecting and placing ads for a target page, i.e., *efficiency evaluation*; (2) effectiveness of position-wise contextual advertising over generic pages, i.e., *effectiveness evaluation*; and (3) effectiveness of position-wise contextual advertising over ambiguous pages, i.e., *ambiguous evaluation*.

5.1. Data and methodology

We conducted experiments to evaluate our approach by using a dataset that contains 50 generic pages, 27 ambiguous pages and 10 224 textual ads. The pages were downloaded from the Internet, with care being taken to ensure that there was an even representation of various areas such as business, electronics, entertainment, health, etc. Then, each page was processed by page segmentation and put through the process of tokenization, stemming and non-keyword filtering (see Section 4.1 for detail). The ad candidates were obtained by querying the Google search engine for nouns and collecting the sponsored ads embedded in the returned pages, supplied by Google AdWords.[3] Then, the candidate ads were also put through the same process as the pages, i.e., tokenization, stemming and non-keyword filtering.

In addition, we selected more than 260 000 articles and 12 000 categories from the wikipedia archives. First, we downloaded the compressed XML file[4] and imported it into a MySQL database by using an XML extraction tool,[5] and then selected articles and categories from the MySQL database. Last, these articles were also put through the process of tokenization, stemming and non-keyword filtering.

In our experiments, we manually judged the content relevance of embedded ads to their placed pages or positions. For each web page, we collected human judgment scores that describe the relevance of ads selected by each of the candidate strategies (presented in Section 5.2). The human judgment scores for the relevance of embedded ads to a page (or a position) were determined by using a scoring method similar to that mentioned in [17,8,18], and were completed by at least two human assessors on a scale between 0.0 and 1.0. The detailed scoring grade is given as follows:

1. *Relevant* (1.0), if the embedded ad is semantically directly related to the main subject of the page (or the ad insertion

[3] Google AdWords, the Google's main advertising product—http://adwords.google.com

[4] Dump—http://download.wikimedia.org/enwiki/20091103/enwiki-20091103-pages-articles.xml.bz2

[5] Mwdumper—http://www.mediawiki.org/wiki/Mwdumper

position). For example, if the ad is about "Peking Hotel", and the insertion position is about "travel in Beijing", it would be scored as 1.0.
2. *Somewhat relevant* (0.5), if the embedded ad is semantically related to the secondary subject of the page (or the position), or it is related to the topic of the page (or the position) in a general way. In the above example, if the page is about "travel in China", the ad would be judged as 0.5 (i.e., the ad is considered to be relevant to its position but somewhat relevant to its page).
3. *Irrelevant* (0.0), if the embedded ad is definitely unrelated to the page (or the position). For example, if the ad is about "Puma shoes" and the page is about "Puma, an American feline", it would be scored as 0.0.

We invited 32 undergraduate students from the Department of Computer Science, all of whom had sufficient Internet browsing experience and judgment ability to conduct the evaluation, to act as assessors to score the embedded ads based on the relevance of each ad to the page and the position. Each embedded ad was first scored by assessors independently, and then, to determine the final relevance score of the ad to the target page and the position, we averaged the relevance scores given by two assessors for each ad.

5.2. Candidate strategies

In our experiments, we compared the performance of the following six contextual advertising strategies in terms of efficiency and effectiveness:

1. *GT*: *global textual relevance* (i.e., we set $w_g = 1.0$ in Eq. (1); and we set $m_k = 1.0$ in Eqs. (5)–(7)). This is similar to conventional contextual advertising approach that adopts keyword matching alone without considering any semantic information, in which the ads are globally selected according to the generic content of the whole page. This is used as the baseline for comparison.
2. *GLT*: *global and local textual relevance* (i.e., we set $w_g = 0.4$; and $m_k = 1.0$). The ads are selected by considering the global and local textual relevance comprehensively. However, the global and local relevance is measured by using the textual similarity alone while neglecting the semantic similarity.
3. *GS*: *global context relevance with semantics* (i.e., $w_g = 1.0$; and $m_c = 0.33$ and $m_e = 0.33$). The ads are selected based on the global context relevance to the target page, without considering the local context relevance to the insertion positions, where the global context relevance is measured by the overall textual and semantic similarity between ads and pages.
4. *GLS*: *global and local context relevance with semantics* (i.e., $w_g = 0.4$; and $m_c = 0.33$ and $m_e = 0.33$). The global and local context relevance are linearly combined together for ad selection and placement. This is a recommended strategy in our approach.
5. *GW*: *global wikipedia relevance*. This is an application of wikipedia matching into contextual advertising, i.e., the work given in [4], where the relevance of ads to pages is measured by wikipedia matching, while the positional placement of ads on pages is not considered. Moreover, to conduct wikipedia matching, we used a set of about 3000 featured articles[6] as the reference model.
6. *GIW*: *global improved wikipedia relevance*. This is an improved wikipedia matching approach to contextual advertising, i.e., the work given in [18], in which the positional placement of ads on pages is also neglected. In this candidate, we used about 260 000 articles as the intermediate reference model, and chose the first selective matching strategy to conduct wikipedia matching.

[6] Wikipedia feature articles—http://en.wikipedia.org/wiki/Wikipedia:Featured_articles

It can seen from the above that the first four strategy candidates are generated according to different settings of the parameters in Eqs. (1) and (5)–(7). In our approach, since the similarity between feature vectors of the same type is measured uniformly by the cosine metric, the parameters (m_k, m_c, m_e) can be trained by cross-validation experiments. In our experiments, the weights in GS and GLS were set to be equal for simplicity.

5.3. Efficiency evaluation

In the first group of experiments, we aimed to evaluate the efficiency of our approach, i.e., the time spent selecting ads which are globally and locally relevant to a target page. The hardware and software setting is shown in Table 1. In our experiments, the work of feature representation for all the ads has been completed offline, i.e., each ad a_j in the ad candidate database $\mathcal{A}$ has been in advance represented and stored as feature vectors. Therefore, in each ad selection for a target page, we only focus on the execution time consumed by page content extraction, page feature representation and page-ads matching. In addition, to improve running performance, we used a similar memory-based buffer to that mentioned in [18] in advance to compress and store all the articles and categories. The experimental results are shown in Table 2.

As we can see from the table, the strategies: GS and GLS, which need to generate additional concept and category vectors on the basis of keyword vectors for a target page, incurs a higher time cost than the benchmark strategies: GT and GLT that use keyword matching alone, but such a time increase is not too serious (less than 800 ms). However, the GIW strategy (i.e., selective wikipedia matching) needs to spend more than 1 s for each ad selection, which is worse in terms of efficiency than our recommended GLS strategy. As for the GW strategy (i.e., wikipedia matching) that needs to match all the reference articles for a target page, its less time spent on one ad selection (about 540 ms) is due to using a very small dataset that comprises about 3000 articles, whereas it needs to take about 1 min to perform this strategy over our dataset, that consists of up to hundreds of thousands of reference articles. Thus, we conclude that our approach obtains a better running performance than other existing ones (the strategies: GW and GIW).

In addition, it also can be seen that compared to the GS, the recommended GLS strategy needs some additional expense, resulting in a decrease about 10% on running efficiency, which is due to that the GLS strategy takes into consideration not only the global context relevance between pages and ad candidates but also the local context relevance, resulting in the need to some additional operations such as page segmentation and insertion position detection. There is a similar situation between the strategies GT and GLT.

5.4. Effectiveness evaluation

In the second group of experiments, we evaluated the advertising accuracy of our approach, i.e., the relevance scores of embedded ads to their pages. In our experiments, we first used the six advertising strategy candidates, respectively, to select the top-N ads and embed them into the page, and then invited human

Table 1
The hardware and software setting.

Item	Explanation
Operating system	MS Windows XP SP2 Professional
CPU	Intel Core (TM) 2 Duo @ 2.93 GHz 2.93 GHz
Running memory	2 GB
Hard disk	500 GB

Table 2
The running time consumed by the six contextual advertising strategy candidates to select relevant ads for a target page.

Strategy	**GT**	**GLT**	**GS**	**GLS**	**GW**	**GIW**
Time (ms)	50	140	785	850	540	1165

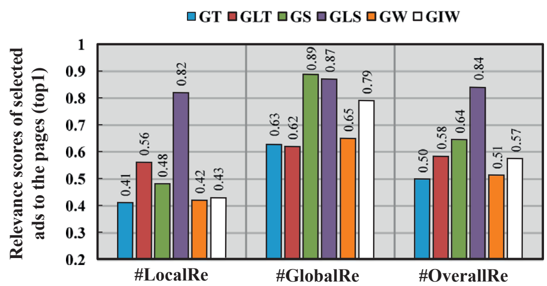


Fig. 4. The average relevance scores for the top-1 ad over general pages.

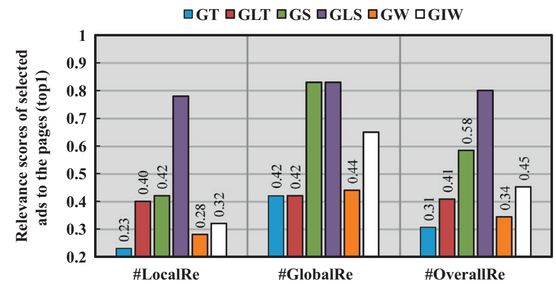


Fig. 6. The average relevance scores for the top-1 ad over ambiguous pages.

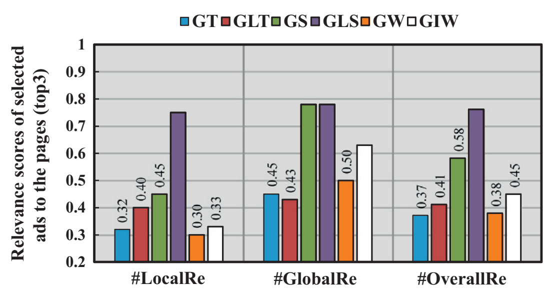


Fig. 5. The average relevance scores for the top-3 ads over general pages.

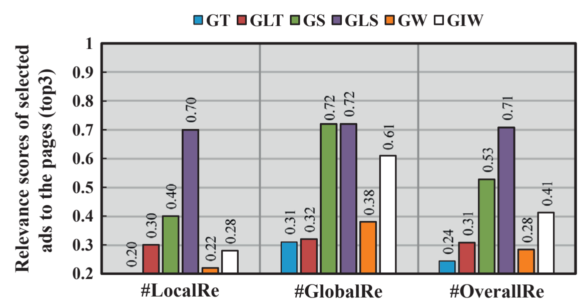


Fig. 7. The average relevance scores for the top-3 ads over ambiguous pages.

assessors to mark score for each ad, based on the relevance of the ad to the insertion position or the page. Last, we averaged the relevance scores given by the human assessors. In our experiments, the advertising accuracy was evaluated from the following three aspects: (1) the local relevance scores of embedded ads to the insertion positions, i.e., *local relevance*; (2) the global relevance scores of embedded ads to the pages, i.e., *global relevance*; and (3) the overall global and local relevance scores of ads to the pages, i.e., *overall relevance*. Below, we describe how to calculate the relevance score for each ad to the page.

Let $\mathcal{H}$ be the human assessors, $\mathcal{Q}$ a set of target web pages, and N the number of ads expected to be embedded to a page. Let $\mathbf{score}(u, a_j, \boldsymbol{p})$ be the relevance score given by the human assessor u for the ad a_j to the page $\boldsymbol{p}$, and $\mathbf{score}(u, a_j, b_i)$ be the relevance score for the a_j to the position b_i. Then, the advertising accuracy for the strategy R (it may be GT, GLT, GS, GLS, GW or GIW) can be measured as follows (where $a_j \in p$ denotes each ad a_j being embedded into the page p, and b_j denotes the insertion position corresponding to the ad a_j):

$$\#\mathbf{LocalRe}(R) = \frac{\sum_{p \in \mathcal{Q}} \sum_{a_j \in p} \sum_{u \in \mathcal{H}} \mathbf{score}(u, a_j, b_j)}{\|\mathcal{Q}\| \cdot N \cdot \|\mathcal{H}\|}$$

$$\#\mathbf{GlobalRe}(R) = \frac{\sum_{p \in \mathcal{Q}} \sum_{a_j \in p} \sum_{u \in \mathcal{H}} \mathbf{score}(u, a_j, p)}{\|\mathcal{Q}\| \cdot N \cdot \|\mathcal{H}\|}$$

$$\#\mathbf{OverallRe}(R) = m_g \cdot \#\mathbf{GlobalRe}(R) + m_l \cdot \#\mathbf{LocalRe}(R)$$

As the purpose of advertising is to select the top-N relevant ads for a given page, we evaluated the advertising accuracy for the top-1 and the top-3 (i.e., $N=1$ and $N=3$). The experimental results are shown in Figs. 4 and 5. It can be seen from the two figures that the four strategies: GT, GS, GW and GIW, perform worse in terms of local relevance than the other two strategies: GLT and GLS. This is due to that the front four strategies only take into consideration how to select globally relevant ads for a page, without considering the placement for the ads on the page (i.e., the ads would be always embedded into a predefined position), consequently, making the embedded ads locally irrelevant to their surrounding page segments. Moreover, the recommended GLS strategy, due to considering the extra local context relevance between ads and their insertion positions, outperforms all the other strategies in terms of local relevance, especially the benchmark strategy (GT): compared to the GT strategy, the average local relevance scores of the GLS strategy dramatically increase by more than 100%, achieving up to about 50% improvement. This also leads to the improvement on the overall advertising accuracy: compared to the other strategies: GT, GLT, GS, GW and GIW, the overall ad-relevance scores of the GLS strategy increase by about 30–100%.

It also can be seen from the figures that the two strategies: GS and GLS, which combine semantic feature information along with keyword matching to measure the context relevance between pages and ad candidates, perform better in the measure accuracy than the other strategies: GT and GLT, which use keyword matching alone: compared to GS and GLS, the overall ad-relevance scores of the formers increase by about 28–95%. By combining wikipedia matching along with keyword matching, the GIW strategy can improve the measure accuracy of context relevance: compared to the benchmark strategy, the global relevance scores of the GIW strategy increase by about 25–45%. However, this strategy takes no account of the local relevance, resulting in not remarkable improvement on the overall accuracy of advertising: compared to the benchmark strategy, the overall ad relevance scores of this strategy only increase by about 10%. Besides, compared to the benchmark, the GW strategy performs fairly in terms of advertising accuracy. This is because in our experiments, we used a small dataset consisting of several thousands of articles as the intermediate reference model, leading to the problem of limited coverage of semantic concepts (see [18] for detail).

From the above, we conclude that our approach, which conducts page-ads matching according to not only the global context relevance but also the local context relevance, can effectively improve the overall global and local accuracy of selected ads to their web pages, consequently, ensuring the effectiveness of position-wise contextual advertising.

5.5. Ambiguous evaluation

In the third group of experiments, to demonstrate that our proposed relevance measure approach (i.e., leveraging the informative wikipedia knowledge along with keyword matching) helps to overcome the problems of homonymy and polysemy, etc. (mentioned in the Introduction section), which cannot be well solved by conventional keyword matching, we have chosen a special dataset which consists of 27 ambiguous pages. In the pages, there are many ambiguous keywords, such as Puma (company vs lion), Rock (person vs music), Driver (software vs car), Game (software vs sports) and Window (OS vs glass). The experimental results about performing the six candidate advertising strategies over the set of ambiguous pages are shown in Figs. 6 and 7.

As we can see from the two figures, the strategies: GS and GLS, again outperform the other two strategies: GT and GLT, which simply use keyword matching, in terms of averaged relevance scores on both the sets of general pages and of ambiguous pages: the overall relevance scores from the GLT strategy reach up to 0.8 for the top-1 and to 0.71 for the top-3, resulting in a significant improvement (more than 100%) over the benchmark strategy. By the above, it is reflected out that the semantic feature information contained in pages (or ads) have better stability than the surface textual feature information, i.e., based on the semantic information, the context relevance between pages and ads can be measured more accurately. It also can be seen from the two figures that whether over the ambiguous pages or over the general pages, the recommended strategy performs better in the overall advertising accuracy than the other two strategies: GW and GIW, which take no account of local context relevance.

From the above, we conclude that by using our proposed context relevance measure approach, i.e., by combining the wikipedia concept and category information along with the textual information to enhance the feature representation for pages and ad candidates, the negative effect caused by the problem of semantic ambiguity can be well degraded.

6. Discussion

Now, contextual advertising has become one of the most important economic engines behind a large number of non-transactional sites. One of the main success factors for contextual ads is their relevance to the surrounding content [2]. However, most existing studies only focus on the selection for relevant ads, without considering the positional effect of the ad placement on the page, consequently, making that, sometimes, the ads are globally relevant to the entire page, but locally irrelevant to the surrounding page segments.

Therefore, the main aim of this paper is to address the framework of our proposal on position-wise contextual advertising by considering global and local context relevance comprehensively for ad selection and placement, so that the selected ads yield contextual relevance to both the entire target page and the ad placement positions. Furthermore, under the framework, to improve the measure accuracy of global and local context relevance, this paper also presents a context relevance measure approach, in which keyword vectors generated by conventional keyword matching techniques are used to measure the textual similarity, and concept vectors and category vectors generated based on the wikipedia knowledge are used to measure the semantic similarity.

In summary, the main advantages of our work include the following three aspects. First, compared to existing wikipedia matching techniques, most of which are not designed for contextual advertising, our approach is more concerned about the running efficiency. In our approach, we significantly decrease the number of relatively time-consuming fulltext matching operations, as well as the depth of graph traversal (when concept mapping and category mapping), resulting in a relatively satisfactory running performance (the time spent on position-wise contextual advertising is generally less than 1 s), thereby, making our approach more applicable for real contextual advertising. Second, our approach obtains better overall advertising accuracy by considering the global and local context relevance comprehensively, so that the embedded ads would be globally relevant to the whole page and locally relevant to their insertion positions. Third, our approach obtains better stability over page-ads matching. This is since we use the wikipedia knowledge to enhance the semantic representation of pages and candidate ads, which, compared to the surface text information, has better stability, i.e., the context relevance between pages and ads can be reflected out more accurately.

In short, this paper presents a beneficial attempt for the problem of position-wise contextual advertising. Compared to existing approaches, the approach proposed in this paper obtains a better balance between advertising accuracy and running efficiency. However, there are still a few factors needed to consider in real application deployment. The first problem is the impact of multiple factors on ad selection. Besides the main factor of the lexical and syntactic contextual relevance, some human cognitive features also have an important influence on the user's ad-click rate. For example, aesthetic factor also plays an important role for online advertising. Will the stylistic factors impact the user's ad-click interest? What is the influential degree of such impact? Which factor is more important? Despite that this study is undoubtedly important in dealing with advertising, it is out of main interest of this paper. We aim to address this in future work.

The second problem is the shortage of our approach itself in real application. In our context relevance measure approach, the basic idea is to use wikipedia concepts and categories to capture the semantic feature information contained in pages and ads. However, the concepts and categories belong to lower level semantics, by which, it is hard to reflect higher level semantic features (e.g., stylistic aesthetic patterns) that generally have better stability than lower level features, thereby, reducing the accuracy on semantic similarity computation at a certain degree. In addition, the running efficiency of our approach outperforms some existing ones, but it still cannot well satisfy the real requirement of in-time contextual advertising. As future work, we therefore plan to research more accurate and efficient context relevance measure approach by capturing higher level semantic feature information implied in texts.

7. Conclusion

In this paper, we have proposed a position-wise contextual advertising approach. On the one hand, we presented an overall framework by considering global and local context relevance comprehensively for ad selection, so that the selected ads yield contextual relevance to both the whole page and the ad insertion positions. On the other hand, we presented a context relevance measure approach by using wikipedia (one of the largest human knowledge bases) to enrich the feature representation for pages and ads, so as to improve the measure accuracy of global and local context relevance. Last, we have also performed experimental evaluation in terms of advertising accuracy and running efficiency over a set of pages and textual ads downloaded from the Internet. The experimental results have demonstrated the effectiveness of our approach.

Acknowledgments

The work is supported by Grants from the NSFC projects (Nos.61202171,61272018,60775037,71090401/71090400), the China Postdoctoral Science Foundation funded projects (Nos.2012M521251 and 2013T60623), the Doctoral Program of Higher Education of China (Nos.20093402110017,20113402110024) and the Zhejiang Provincial Natural Science Foundation of China (Nos.Q13F020034,Y1100769, Y1090328).

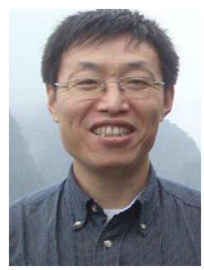
References

[1] A. Anagnostopoulos, A. Broder, E. Gabrilovich, V. Josifovski, L. Riedel, Just-in-time contextual advertising, in: CIKM'07, ACM, New York, NY, 2007, pp. 331–340.
[2] A. Broder, M. Fontoura, V. Josifovski, L. Riedel, A semantic approach to contextual advertising, in: SIGIR'07, ACM, New York, NY, 2007, pp. 559–566.
[3] Search engine watch, ⟨http://searchenginewatch.com/2183531⟩, 2003.
[4] A.N. Pak, C.-W. Chung, A wikipedia matching approach to contextual advertising, WWWJ 13 (3) (2010) 251–274.
[5] A.N. Pak, Using wikipedia to improve prevision of contextual advertising, in: Proceedings of the 4th International Conference on Human Language Technology: Challenges for Computer Science and Linguistics (LTC'09), Springer-Verlag, Berlin, Heidelberg, 2011, pp. 533–543.
[6] T.-K. Fan, C.-H. Chang, Sentiment-oriented contextual advertising, KAIS 23 (3) (2010) 321–344.
[7] P. Chatterjee, D.L. Hoffman, T.P. Novak, Modeling the clickstream: implications for web-based advertising efforts, Mark. Sci. 22 (4) (2003) 520–541.
[8] T. Mei, X. S. Hua, S. P. Li, Contextual in-image advertising, in: MM'08, ACM, New York, NY, 2008, pp. 439–448.
[9] T. Mei, X.S. Hua, S.P. Li, Contextual internet multimedia advertising, Proceedings of the IEEE 98 (8) (2011) 1416–1433.
[10] T. Mei, L.S. Li, X.S. Hua, S.P. Li, Imagesense: towards contextual image advertising, ACM TOMCCAP 8 (1) (2012).
[11] G. Salton, A. Wong, C. Yang, A vector space model for automatic indexing, Commun. ACM 18 (11) (1976) 613–620.
[12] S. Gupta, G.E. Kaiser, P. Grimm, M.F. Chiang, J. Starren, Automating content extraction of html documents, WWWJ 8 (2) (2005) 179–224.
[13] B. Ribeiro-Neto, M. Cristo, P.B. Golgher, E.S. Moura, Impedance coupling in content-targeted advertising, in: SIGIR'05, ACM, New York, NY, 2005, pp. 496–503.
[14] V. Murdock, M. Ciaramita, V. Plachouras, A noisy-channel approach to contextual advertising, in: SIGKDD workshops 07, ACM, New York, NY, 2007, pp. 21–27.
[15] S. Papadopoulos, F. Menemenis, Y. Kompatsiaris, B. Brato, Lexical graphs for improved contextual ad recommendation, in: ECIR'09, 2009, pp. 216–227.
[16] S. Gupta, G.E. Kaiser, P. Grimm, M.F. Chiang, J. Starren, Knowledge distribution via shared context between blog-based knowledge management systems: a case study of collaborative tagging, ESWA 36 (2) (2009) 10627–10633.
[17] J. Hu, L.J. Fang, Y. Cao, H.-J. Zeng, H. Li, Q. Yang, Z. Chen, Enhancing text clustering by leveraging wikipedia semantics, in: SIGIR'08, ACM, New York, NY, 2008, pp. 179–186.
[18] Z.D. Wu, G.D. Xu, Y.C. Zhang, P. Dolog, C.L. Lu, An improved contextual advertising matching approach based on wikipedia knowledge, COMPJ 55 (3) (2012) 277–292.
[19] Z.D. Wu, G.D. Xu, Y.C. Zhang, R. Pan, Z.W. Hu, J.F. Lu, Leveraging wikipedia concept and category information to enhance contextual advertising, in: CIKM'11, ACM, New York, NY, 2011, pp. 2105–2108.
[20] X.H. Hu, X.D. Zhang, C.M. Lu, E.K. Park, X.H. Zhou, Exploiting wikipedia as external knowledge for document clustering, in: SIGKDD'09, ACM, New York, NY, 2009, pp. 389–396.
[21] J. Hu, G. Wang, F. Lochovsky, J.-T. Sun, Z. Chen, Understanding user's query intent with wikipedia, in: WWW'09, ACM, New York, NY, 2009, pp. 471–480.
[22] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, Cambridge, MA, 2004.
[23] Comprehensive perl archive network, ⟨http://search.cpan.org/~jzhang/HTML-ContentExtractor-0.03/lib/HTML/ContentExtractor.pm⟩, 2007.
[24] H.C. Wu, R.W.P. Luk, K.F. Wong, K.L. Kwok, Interpreting tf-idf term weights as making relevance decisions, ACM TOIS 26 (3) (2008) 13–50.

Zongda Wu is an associate professor in Computer Science at Wenzhou University. He received his B.Sc. degree in Computer Science from Wenzhou University, in 2005 and Ph.D. degree in Computer Science from Huazhong University of Science and Technology (HUST), in 2009. Now, he is also a postdoctoral research fellow with School of Computer Science and Technology at University of Science and Technology of China (USTC). His research interests are primarily in the area of computational advertising, information retrieval and query processing.

Guandong Xu is a research fellow in Advanced Analytics Institute, Faculty of Engineering and Information Technology, University of Technology, Sydney. He obtained his Ph.D. degree in Computer Science from Victoria University, in 2008. After that he worked as a postdoctoral research fellow in Center for Applied Informatics at Victoria University and then postdoc in Department of Computer Science at Aalborg University, Denmark. He is an endeavor postdoctoral research fellow in the University of Tokyo, in 2008. His research interests include web information retrieval, web mining, web services, etc.

Chenglang Lu is a lab assistant in Computer Science at Wenzhou University. He received his B.Sc. degree in Computer Science from Wenzhou University, in 2005 and Master of Software Engineering degree from Huazhong University of Science and Technology (HUST), in 2009. Now, he is currently pursuing his Ph.D. degree from School of Marine Engineering, Northwestern Polytechnical University, China. His research interests include content-based video information retrieval and query processing.



Enhong Chen is a professor and a vice dean at School of Computer Science and Technology, University of Science and Technology of China (USTC), IEEE senior member. He received his Ph.D. degree in Computer Science from USTC, in 1996. He has been actively involved in the research community by serving as a PC member for more than 30 conferences, such as ICTAI 2006, ICTAI 2007, AIRS 2009, AIRS 2010 and KDD 2010. His research interests include semantic web, machine learning and data mining, web information processing, constraint satisfaction problem, etc.

Yanchun Zhang is a professor and director of Centre for Applied Informatics at Victoria University. He obtained his Ph.D. degree in Computer Science from the University of Queensland, in 1991. He has been active in areas of database and information systems, web and internet technologies, web mining, web services, and e-health. He has published over 200 research papers in international journals and conference proceedings. He was a member of Australian Research Council (ARC) College of Experts (2008–2010). He is the Editor-In-Chief of World Wide Web journal and Health Information Science and Systems journal, and Chairman of Web Information Systems Engineering (WISE) Society.

Hong Zhang is an engineer in Information Management at Institute of Science and Technology Information of Zhejiang Province. He received his B.Sc. degree in Management Science from Ningbo University, in 2003 and his Master of Software Engineering degree from Hangzhou Dianzi University, in 2011. His research interests include rural science and technology information research, and competitive intelligence research.