

Observing Facial Expressions and Gaze Positions for Personalized Webpage Recommendation

Songhua Xu*

Oak Ridge
National Laboratory,
Oak Ridge, Tennessee 37831,
USA

Hao Jiang*

Department of
Computer Science,
The University of Hong Kong
Hong Kong S.A.R., P.R. China

Francis C.M. Lau

Department of
Computer Science,
The University of Hong Kong
Hong Kong S.A.R., P.R. China

ABSTRACT

We propose a new method for personalized webpage recommendation. The method is capable of inferring a user's personal reading interest distribution according to implicit user feedbacks coming from the user's past online reading activities. With the inferred user reading interest distribution, we can recommend webpages in a search result set to a user in a personalized way. Our method is featured by its novel approach to observe the facial expressions and gaze positions of a user during the user's online reading activities as two types of implicit user feedbacks for estimating the user's reading interest distribution. To capture these implicit user feedbacks, we use an ordinary web camera and a customized web browser in the setup. The setup allows us to measure the distribution of the reading time a user spends in his or her reading activities over materials of different contents. With all the captured information, our method then estimates a user's reading interest distribution by finding correlations between the implicit feedbacks of a user with the contents of the read materials. Given the estimated user reading interest distribution, our algorithm can further predict the user's potential reading interest in any new webpage. Consequently, our algorithm can produce a personalized webpage recommendation for all the result webpages in an online search session. We compared the performance of our method with that of several mainstream commercial search engines as well as a recent personalized webpage ranking algorithm. The comparison results clearly show the superiority of our new method for personalized webpage recommendation.

Categories and Subject Descriptors

H.1.2 [User/Machine Systems]: Human Factors, Human Information Processing; H.3.5 [Online Information Services]: Commercial Services, Web-based Services; H.5.2 [User

*To be considered as equal first authors.

Interfaces]: Input Devices and Strategies, Interaction Styles; I.2.1 [Artificial Intelligence]: Applications and Expert Systems; I.2.7 [Natural Language Processing]: Text Analysis

General Terms

Algorithms, Design, Experimentation, Human Factors, Measurement, Performance

Keywords

Web content recommendation, personalized recommendation, facial expression, eye-tracking, implicit user feedback

1. INTRODUCTION

Recently, researchers have shown increasing interests in improving user experiences when building advanced information retrieval and web search systems. Among the popular topics under this trend, personalized webpage recommendation and re-ranking of webpages are particularly heated because of their critical impact on users' web search experiences. A personalized webpage recommendation method or system needs some kind of user feedback in order to determine a user's web search preferences. However, counting on explicit user feedbacks is not always feasible, and even if they are available they tend to be too costly. And so some researchers turn to implicit user feedbacks, such as browsing history and document reading time.

In this paper, we propose a personalized webpage recommendation algorithm based on a user's reading interest distribution. Such a personal reading interest distribution is constructed according to implicit user feedbacks derived from the user's previous web browsing behaviors. The user's behavior data are acquired using a web camera and a customized web browsing setup. In particular, we use facial expression and gaze position to infer the user's reading interests. Using such implicit user feedback data we capture, we correlate the inferred user reading interests with the contents of webpages that the user has previously browsed. Through this correlation, we try to derive the user's reading interest distribution. Given the estimated reading interest distribution of a user, for every new search result webpage, we can predict the potential reading interest of the user in the webpage. We apply the predicted user reading interest to all the search result webpages to produce a personalized webpage recommendation for the user.

The remainder of this paper is organized as follows. We first survey some related work in Sec. 2. We then briefly describe how to detect a user's gaze positions and facial expressions using a web camera in Sec. 3. After that, we discuss how to measure word-level user reading interests according to the captured gaze positions and facial expressions of the user during his or her previous webpage reading or browsing activities in Sec. 4. According to the measured word-level reading interest of a user, we introduce a method to predict the potential reading interest of a user with respect to a search result webpage, with which we can further produce a personalized webpage recommendation for the user. Sec. 5 discusses such a process. In Sec. 6, we report some sample results of our algorithm for personalized webpage recommendation. To quantitatively evaluate the effectiveness of our method, we compare the performance of our algorithm with that of a few major commercial web search engines, as well as a recent personalized webpage ranking algorithm. Finally, we conclude the paper and point out some future working directions in Sec. 7.

2. RELATED WORK

2.1 Personalized Recommendation

The closest work related to our study is the personalized webpage ranking algorithm proposed in [29]. Their method estimates a user's reading interest based on user attention time which is the time a user spent previously in reading a webpage. Based on the captured user attention time information, their algorithm predicts the potential attention times of the user for unvisited webpages. When applied to a set of search result webpages, their algorithm produces a personalized webpage ranking. Also related is the personalized document, online image and video recommendation algorithm proposed in [27], which adopts a commodity eye-tracking based setup to capture the user attention time for online content recommendation. In both of the above methods, the authors only utilized the user attention time information when generating personalized webpage recommendations. Our study in this paper additionally explores the facial expression of a user to more accurately infer the reading interest of the user, and to make more accurate personal webpage recommendation. Besides facial expressions, we also use a head pose based method to estimate the gaze position of a user during browsing. We believe our approach works more effectively and reliably than the commodity eye-tracking based approach introduced in [27]. Later in this paper, we will quantitatively demonstrate the superiority of our method to the algorithm proposed in [27] through experimentation.

Sieg et al. [22] proposed a method that can derive user interest scores for all the concepts in an ontology. Their method employs a spreading activation algorithm to update user interest scores based on a user's continuous information access behaviors. Qiu and Cho [20] introduced an automatic algorithm to generate personalized search results by detecting and classifying the interests of different users according to their online behaviors. The study conducted by Dou et al. [6] applied five personalized search methods over a 12-day MSN query log. The authors found that personalized searches cannot perform significantly better than traditional web searches in some situations. They further pointed out that click entropy provides a good criterion for

whether a query should be executed in a personalized way. Chirita et al. [5] introduced a method which modifies a query by adding terms extracted from a personal information repository to perform personalized web searches. The CubeSVD method suggested by Sun et al. [24] employed a singular value decomposition approach to analyze the click-through data for discovering the relationships between a user's submitted queries and the corresponding search result webpages. The captured relationships are then used to build personalized web searches. In our previous work [28], we proposed an eye-tracking based document summarization algorithm, which can produce personalized document summaries according to the reading time information gathered from a user's previous reading activities.

2.2 Facial Expressions as a Type of Implicit User Feedback

There is a long history of research endeavors on using human facial expressions for building intelligent systems. A lot of them took place in the medical domain, e.g., to detect Amygdala response through facial expressions of children and adults [13], to capture attention through emotions detected from facial expressions of disabled patients [25], etc. The computer science field is mostly concerned with developing novel facial expression detection algorithms. Many researchers have proposed image processing algorithms to detect or track human faces from images [18]. Equally well studied is the problem of face recognition, due to its wide applicability such as in identity verification [30]. However, to the best of our knowledge, we are not aware of any work that has explored user facial expressions as a type of implicit user feedback for building personalized information management or recommender systems.

As a specialized line of research, eye-tracking is recognized as a promising approach for obtaining implicit user feedbacks [4]. There exist both commercial eye-trackers such as those electro-oculographic (EOG) based systems [2, 3] and web camera based commodity approaches [21, 14, 32] that try to capture user gaze positions. A professional eye-tracking device however is very expensive, which severely hinders its adoption. The alternative is the web camera (webcam) has become more or less a default device of any new laptop. Given the huge cost advantage of the webcam, more researchers have turned to it to develop eye-tracking algorithms and systems [23, 8]. Likewise in this paper, we adopt a web camera based approach to capture user facial expressions and gaze positions for understanding a user's personal reading interest distribution in making personalized webpage recommendation.

3. DETECTING GAZE POSITIONS AND FACIAL EXPRESSIONS OF A USER

3.1 Detecting User Gaze Positions

To determine the gaze positions of a user during the user's online reading activities, we first identify the orientation of the user's face. As mentioned earlier, our approach uses a web camera as the basic input device, which is coupled with computer vision techniques to track a user's eye movement. We did not choose a commercial eye-tracker, e.g., a wearable EOG system [2, 3], mainly because of its very high cost. We will later show in the experiment section of this paper that

the level of accuracy achieved using our web camera based setup can already obtain very satisfying results in making personalized webpage recommendation.

Our system detects a user's gaze point by using a mouse replacement software, called "Enable Viacam" [15]. Enable Viacam was originally developed to allow handicapped people to control the mouse through their head poses. The software first detects the orientation of a user's face, and then predicts the screen region that the user is looking at. With the predicted screen region, the software then moves the mouse to the gaze point of the user. In our work, we detect the gaze positions of a user through reading the intermediate results output by the software. In our experiments, on a 19-inch computer screen, the error of the detected gaze point is from 2 to 4 centimeters, depending on the head-moving behavior of a user. The error is also affected by the size of the screen. For example, on a 24-inch screen, typical errors of gaze point detection would be about 3 to 6 centimeters horizontally and 2 to 3 centimeters vertically.

In the personalized webpage recommendation method proposed in [27], the authors also employed a vision-based eye-tracking approach [32] to capture the gaze points of a user. Their approach could achieve a smaller gaze detection error, in the range of 1 to 2 centimeters for a 19 inch computer screen. However, their method has the unreasonable assumption that once the initial calibration process is done, users will not move their heads while browsing the web. In contrast, the gaze position detection in our work here does not require any initial calibration, and would allow users to freely move their heads during operation. Also, as mentioned earlier, a higher accuracy in gaze detection is not most necessary for our application, which we will further explain in Section 6 when we present our experimental results.

3.2 Detecting User Facial Expressions

To detect user facial expressions during online reading or browsing activities, we adopt a recent facial emotion analyzer "eMotion" [1]. The output of eMotion is a series of probabilistic values that represent the likelihood of the user making certain facial expressions. In our approach, we use eMotion's facial expression detection output to derive the reading concentration of a user. We introduce the concept of reading concentration for measuring the degree of attentiveness and reaction during webpage reading.

Let $rc(u, t)$ be the *reading concentration sample* of the user u at time moment t . For a specific user u , the larger the reading concentration value, the more attentively the user reads a webpage, which means that the contents of the webpage attract the user more. In our system, we measure a user's reading concentration at any moment according to the output from the facial emotion analyzer. More concretely, $rc(u, t)$ is estimated as follows:

$$rc(u, t) = 1 - F_{neutral}, \quad (1)$$

where $F_{neutral}$'s value range is $[0, 1]$, which represents the probability of having a neutral facial emotion at the moment t , as detected by the human facial emotion capturing software eMotion [1]. Here we assume if a piece of information appears interesting to a user, then the user will tend to display some non-neutral expressions while reading the

information.

Combining the above reading concentration measurement, derived from user facial expressions, with the gaze points detected using the method described in Sec. 3.1, we can then estimate a user's reading interest distribution. We will look at this in details in the next section.

4. MEASURING WORD-LEVEL USER READING INTERESTS

Given the detected gaze samples of a user, we can obtain a series of two-dimensional points, each representing a gaze position of the user at a certain moment. To link gaze samples with the contents of the user's reading materials, we need to anchor all the gaze samples, along with their corresponding reading concentration samples, onto individual words for deriving word-level user reading interest. The basic assumption here is that a region with higher reading concentration contains more interesting contents to the user.

To measure word-level user reading interest, we first introduce a concept called *the snapshot of a webpage*, which refers to the part of a webpage displayed on the screen at a certain moment. The snapshot will change correspondingly if the user changes the part of the webpage displayed on the screen, e.g., by scrolling up and down, or by clicking to open another new webpage. Every reading concentration sample we acquire using the method introduced in Sec. 3 must be captured with respect to a certain webpage snapshot, as a user can only look at one screen region at a time. Due to this property, we link each reading concentration sample to a few words in the webpage, which are closest to the detected gaze point of the user. To more elaborately assign reading concentration samples to words, we adopt a Gaussian model. Assuming at a certain moment t , the detected gaze point is at position (x^t, y^t) on the screen. For each individual word w_i displayed in the current webpage snapshot, we first compute w_i 's coordinate as the center of the bounding box of the word's screen displaying region. We denote the center position of the displaying region as (x_i, y_i) . Adopting the above notation, for a reading concentration sample $rc(u, t)$ acquired at the moment t from the user u , we assign the user's reading concentration on the word w_i at the moment, $RC(u, w_i, t)$, as follows:

$$RC(u, w_i, t) = rc(u, t) \exp\left(-\frac{(x_i - x^t)^2}{2\sigma_x^2} - \frac{(y_i - y^t)^2}{2\sigma_y^2}\right). \quad (2)$$

The free parameters σ_x and σ_y specify how "diffusively" a user scans words when reading webpages. In our current implementation, we assign values for the two parameters using the average width and height of all the words' displaying bounding boxes in the webpage. Intuitively, when the distance between a word and the user's gaze point decreases, a larger portion of the reading concentration sample will be assigned to the word. All the reading concentration samples will be assigned to individual words following the above procedure. Finally, the overall reading concentration on a word in a webpage is calculated as the sum of all the reading concentration samples assigned to the word. Note that a word occurring multiple times in a webpage is regarded as multiple distinct words in the above process. Therefore, the reading concentration assignments over all the occurrences

of a word will be added together. Overall, the reading concentration over a word w_i of the user u in our method is estimated as follows:

$$RC(u, w_i) = \sum_{j=1}^{z(t)} RC(u, w_i, t_j) \\ = \sum_{j=1}^{z(t)} \left\{ rc(u, t_j) \exp\left(-\frac{(x_i - x^{t_j})^2}{2\sigma_x^2} - \frac{(y_i - y^{t_j})^2}{2\sigma_y^2}\right) \right\}. \quad (3)$$

Here we assume that our system has collected a total of $z(t)$ gaze position samples, along with their corresponding reading concentration samples $rc(u, t_1), \dots, rc(u, t_{z(t)})$. These samples are acquired at the moments $t_1, \dots, t_{z(t)}$ respectively. Recall that the notation (x^{t_j}, y^{t_j}) denotes the coordinate of the user's gaze point at time t_j .

Note that during our above processing, we remove all the stop words in a webpage since they do not provide much insight on the contents of the reading materials and hence could not reflect the user's reading interest. For words in a webpage that are not displayed on the screen, their reading concentration values are represented using a special unspecified tag rather than being assigned to zero. Finally, if a word occurs more than once over all the webpages browsed by a user, we will gather all the user's reading concentration assignments over the word together to derive an overall reading concentration value, which should reflect to some accuracy the user's overall reading interest on the word.

5. PERSONALIZED WEBPAGE RECOMMENDATION

5.1 Predicting User Interest on Unseen Webpages

For a new webpage that the user has not previously encountered, we can predict the user's potential reading interest over the webpage according to the user's reading behaviors on all the previously visited webpages of the user. Here we first carry out several preprocessing steps to remove texts on the webpage that are irrelevant to the webpage's main contents, such as tags, advertisements, website navigation bars, and redirection links [19]. To predict a user's interest over webpages, we first represent every webpage as a vector of distinct words. Such a representation is necessary for establishing the content relatedness between words on two webpages so that we can predict the user interest on a previously unvisited webpage from his or her already visited webpages. For each word on the webpage, we can derive its word frequency as its $TF \times IDF$ coefficient [26]. A webpage can be then represented as a vector of word frequencies.

Based on the above representation, we can then predict the potential user interest on a new webpage as the sum of the user's interests over all the words occurring in the webpage. Here the user interest on a word w_i can be estimated by the sum of the user's reading interest on the same word or similar words that occur in the user's previously visited webpages. For the latter type of user interest, the contribution of a word w'_i to the estimation of reading interest over a related word w_i is calculated as the product of w'_i 's reading concentration and the semantic similarity between w_i and w'_i . In our current implementation, the semantic similarity between a pair of words is computed using WordNet [17]. In WordNet, all the English words are organized as a network,

where the connection edges in the network are established according to words' mutual lexical relations. The semantic similarity between any pair of words can be measured as a real number in the range of $[0, 1]$, where 0 represents two completely unrelated words; while the similarity between two identical words is 1.

5.2 Personalized Webpage Recommendation

Once we have predicted the potential reading interest of a user over every webpage in a web query search result set, we can then produce a personalized webpage ranking over these search result webpages according to the predicted user interests over each search result webpage. However, ranking webpages only according to the potential user interest is not sufficient as other factors such as the relatedness of webpage content to the query keyword as well as the reputation of the webpage's source site should also be considered. These issues are well addressed by existing web search algorithms. To account for these issues, we propose to combine our algorithm with one of the conventional non-learning based document retrieval algorithms. In this work, we choose the topic-sensitive PageRank algorithm [9] which has comprehensively considered all the above factors. Also, like any example based methods, our new algorithm suffers from the cold-start problem, i.e., if there are not enough webpages browsed by the user in the past, there will not be sufficient reading time samples available to train our system. Under that circumstance, the inferred personal reading interest distribution of a user will likely only be able to provide weak and unreliable indications on the user's personal preference over the candidate webpages. Again, our above hybrid approach for webpage ranking can help overcome such a cold-start problem.

Our prototype system implementation mainly consists of three parts: 1) a user reading concentration acquisition component based on our facial emotion detector; 2) personal reading interest understanding based on a user's previously web surfing behaviors; 3) user potential interest prediction for unvisited webpages based on the user's past webpage reading behaviors. To apply our algorithm for re-ranking search result webpages, each time when a web search query is submitted by a user, we first use the Google search engine to obtain the top 300 search result webpages. We then predict the potential reading interest of the user over these 300 webpages, using the aforementioned hybrid version of our method, which has incorporated the topic-sensitive PageRank algorithm [9]. These webpages are then recommended to the user in descending order of predicted user interest.

Taking the merits of both our learning based and the classical document retrieval approaches, we predict an overall user interest term for each webpage D_i as follows:

$$OI_u(D_i) = (1 - \lambda) \cdot \theta(u, D_i) + \lambda \cdot \frac{2 \exp(-\gamma \cdot TR_i)}{1 + \exp(-\gamma \cdot TR_i)}, \quad (4)$$

where D_i is one of the top 300 webpages fetched from a commercial search engine; $OI_u(D_i)$ stands for the user u 's overall interest in reading the webpage D_i ; and $\theta(u, D_i)$ denotes the user u 's potential reading interest over the document D_i as predicted by our method introduced in Sec. 5.1. The user tunable parameter λ ranges from 0 to 1, which moderates the balance point between our learning based and the PageRank

#	Search Keywords	SES_{Google}	SES_{Ours}	Reduction(%)
1	apple	98	165	41%
2	car	136	231	41%
3	barcelona	117	176	34%
4	da vinci	90	201	55%
5	ETS	129	157	18%
6	gnome linux	109	142	23%
7	greenhouse effect	65	95	32%
8	happy new year	129	167	23%
9	NBA	141	186	24%
10	olympics	63	97	35%
11	wow	196	278	29%
12	great wall	164	182	10%
13	hurricane	127	227	44%
14	iron man	113	147	23%
15	moon	132	218	39%
16	national treasure	126	197	36%
17	porsche	65	134	51%
18	forbidden kingdom	132	201	34%
19	tiger	191	272	30%
20	west lake	105	134	22%
Average		121	180	32%

Table 1: Twenty examples of query words and their corresponding search error scores when running our algorithm. We report the search error scores on the rankings produced by Google, SES_{Google} , and by our algorithm, SES_{Ours} respectively. On average, our algorithm achieves a 32% of SES reduction compared to Google’s ranking result. These performance evaluation results are also visually compared in Figure 2.

based webpage ranking considerations. In our experiments, we set λ as follows:

$$\lambda = \exp\left(-\frac{n_u}{100}\right), \quad (5)$$

where n_u is the number of articles the user u has read in the past, whose reading interests are known to our algorithm. TR_i denotes the rank of the document D_i among the top 300 search result webpages. Such a rank is generated by the topic-sensitive PageRank algorithm [9] in our system. Documents with low ranks will be assigned a small value for the TR_i term. The parameter γ specifies how sharply this drop-off is, whose typical value in our experiment is set to 0.2.

6. EXPERIMENTS

Our experiments were conducted in the following setting. Twenty volunteers from our university were invited to participate in our experiments by using our prototype system. We equipped each volunteer’s laptop with a web camera and our customized web browser. All the participants were asked to use our customized web browser throughout the entire experiment period for web surfing. We implemented our customized web browser as a stand-alone program on the Windows XP platform. We use such a customized web browser to capture reading concentration data when a user surfs the Internet. Figure 1 shows the snapshots of our customized web browser running an example query. Given these data, each time a user submits a web search request, our customized web browser will return the search results following

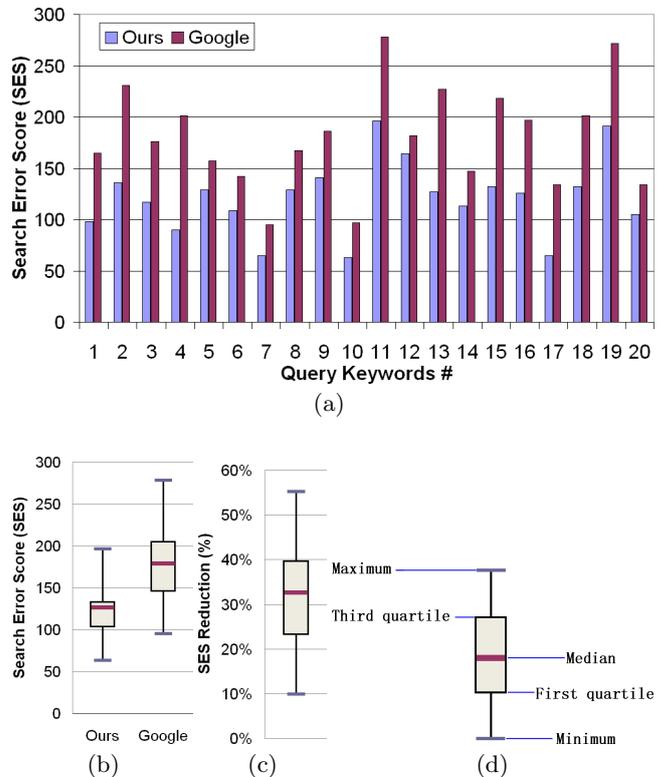
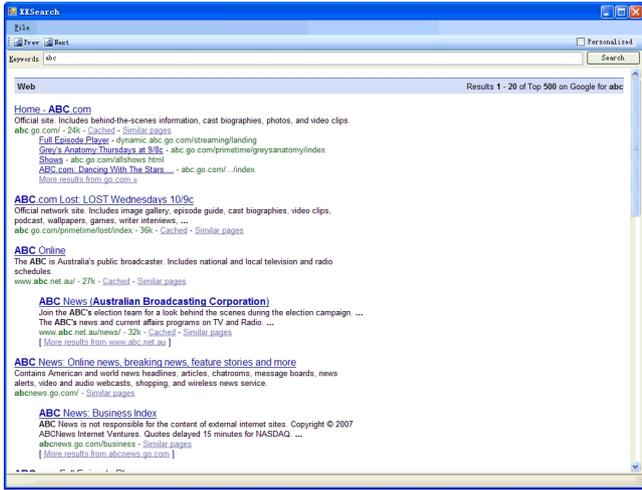


Figure 2: (a) Search error scores (SES) of webpage rankings over top search result webpages for twenty example queries generated by our algorithm and Google respectively. (b) Boxplot diagrams reporting statistics of the SESs of webpage rankings produced by our algorithm and Google for the twenty example queries respectively. (c) Another boxplot diagram that reports statistics of the SES reduction rates by rankings produced by our algorithm with respect to Google’s webpage rankings. (d) illustrates the meanings of key elements used in our boxplot diagram. The corresponding performance comparison data are reported in Table 1.

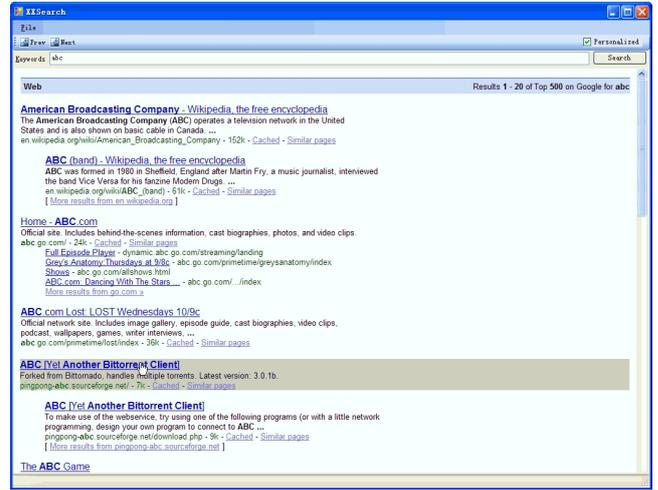
the personalized webpage ranking recommended by our algorithm to the user. In our experiments, we quantitatively evaluate the quality of such personalized webpage rankings by comparing the rankings generated by a few mainstream commercial search engines as well as the rankings produced by a recently proposed personalized web recommendation algorithm [27].

6.1 Ranking Quality Measurement

The traditional precision, recall and F-rate measurements for information retrieval all assume simple binary decisions on the relevance of a search result item to its corresponding query [11]. However, in practice, a webpage can be partially related to a query; and it can also be more closely related to a query than another webpage. Such relative webpage to query relatedness is not captured and measured by the traditional precision, recall and F-rate measurements, nor their variants or extensions such as the cumulative gain-based evaluation metric [12]. In the context of personalized



(a)



(b)

Figure 1: Snapshots of our customized web browser running an example query “abc”. (a) gives the search results retrieved by Google without employing our personalized recommendation algorithm; (b) shows the search results ranked according to the recommendation suggested by our system.

webpage ranking, we can more intuitively understand the necessity for measuring a ranking’s capability to place relatively more important search results at the top of a search result list, as we human users can often intuitively feel that some webpages are more interesting or useful than others in a search result webpage set. Given the fact that in most real-world search scenarios, there usually are a large amount of search results returned for a query, the capability to place those most important items towards the top of a search result list is very critical for the usability of a search engine [10].

Considering the practical need of measuring a ranking’s capability in order to prioritize the link positions of the most important search result webpages, and also inspired by the search quality measurement method in [31], we adopt the following weighted search quality measurement in our experiments for evaluating the quality of a webpage search result ranking:

$$SES(Q) \triangleq \frac{\sum_{i=1}^{N_{top}} \omega_i X_i}{\sum_{i=1}^{N_{top}} \omega_i}, \quad (6)$$

where $SES(Q)$ stands for the *search error score* of the ranking Q over the search result set of a query. $X_i \triangleq |IR_i - i|$ is the difference between the ranking of the i -th search result webpage and the webpage’s rank in a user’s expected ideal ranking, IR_i . N_{top} is the number of top search result webpages which receive most user attention. In all our experiments, N_{top} is assigned to be 20. And ω_i ’s are the position dependent search result weights, which are defined as follows:

$$\omega_i \triangleq \begin{cases} 2.5 & (1 \leq IR_i \leq 5) \\ 2.0 & (6 \leq IR_i \leq 10) \\ 1.5 & (11 \leq IR_i \leq 15) \\ 1.0 & (16 \leq IR_i \leq N_{top}) \end{cases}. \quad (7)$$

These weights modulate the influence of webpages of different ranking positions on the overall quality of a search

result webpage ranking. In our experiments, the ideal webpage ranking of a user is manually specified by the user at the end of the corresponding web search experiment when the user is first asked to read through all the top N_{top} search result webpages and then make a manual ranking over these webpages according to the user’s reading interest and personal preference.

6.2 Experiment Results

To conduct our experiments, all the volunteers were asked to perform web searches as usual for a period of two weeks, but using our personalized webpage recommendation system. The participants are engineering majored students with intensive Internet search experiences. During web searching, the reading concentration data of individual users over each search result document are collected by our system in a non-intrusive manner. In the second week, each user was also asked to perform web search using some given query words. For a query word search, the user was asked to give an ideal ranking of the first N_{top} result webpage set according to his/her personal interests. Such data will be used as the testing data in our experiments. We use the participant’s reading concentration data collected in the first week of our experiment as the training samples. We report our experiment results on twenty example queries in Table 1. These performance evaluation results are also visually compared in Figure 2. In Table 2, we also report search error scores for search results obtained using three popular commercial web search engines—Google, Yahoo! and Bing (www.bing.com). For comparison purposes, we also report the performance of the recently proposed “AT08” algorithm [27], which also has access to users’ eye-tracking data collected in the first week of our web search experiments as training data. Key statistics on the SES reduction ratios with respect to Google’s ranking results are visually compared in Figure 3 using boxplots. All these comparisons clearly show the ability of our algorithm for generating webpage rankings that best resemble a user’s expected ideal

#	Search Keywords	Google	Yahoo		Bing		AT08		Ours	
		SES	SES	Reduction(%)	SES	Reduction(%)	SES	Reduction(%)	SES	Reduction(%)
1	blizzard	76	96	-26%	77	-2%	66	13%	52	31%
2	diving	163	186	-14%	172	-6%	112	31%	83	49%
3	dweep	82	98	-20%	93	-14%	79	3%	40	51%
4	earthquake	174	181	-4%	189	-8%	131	24%	107	38%
5	everest	242	259	-7%	235	3%	173	29%	156	36%
6	eyes on me	63	70	-11%	60	5%	44	30%	38	40%
7	gnome linux	142	169	-19%	136	4%	115	19%	76	47%
8	grand canyon	199	225	-13%	207	-4%	172	14%	138	31%
9	phoenix	278	331	-19%	293	-6%	229	18%	147	47%
10	prison break	88	95	-8%	86	3%	76	14%	50	43%
11	risc	150	157	-5%	148	2%	132	12%	95	37%
12	the beach	107	107	0%	100	7%	101	5%	74	31%
13	tomb raider	73	89	-22%	78	-7%	66	9%	51	31%
14	transformers	86	83	3%	99	-15%	69	19%	51	40%
15	world cup	81	93	-15%	79	2%	59	27%	54	33%
Average		134	149	-12%	137	-2%	108	18%	81	39%

Table 2: Fifteen example search results and their corresponding search error scores (SEs) and SES reduction ratios with respect to Google’s ranking by Yahoo!, Bing, the “AT08” algorithm, and our algorithm respectively. Both the “AT08” algorithm and our algorithm have access to the user data collected from the volunteers in our experiments as the training data. Key statistics of the SES reduction ratios are also visually compared in Figure 3 in the form of boxplots.

webpage rankings among all the peer systems and methods.

6.3 Personalization for Domain-Specific Search

Compared with searches over general information, our personalized webpage recommendation method proposed in this paper should show even greater effectiveness for domain-specific searches for which the users’ personal preference and interest tend to play a more important role in ranking search results. To explore this hypothesis, we also conducted several experiments on domain-specific search scenarios. We chose online shopping as the specific domain, because of its importance in many people’s daily life, and its high market potential. We developed a customized recommender systems for eBay, which implements all the components of our system. Our customized recommender system searches customer-to-customer commodities to generate personalized ranking results for an individual user. In this set of experiments, we did not use the Google search results as the baseline method, but instead used eBay’s search results as the baseline. We report our experiment results on several example queries in Table 3. All the environment settings and measurements are kept the same as our experiment settings described in Sec. 6.2. We also evaluated the performance of our algorithm, Google (searches restricted within the eBay site), and the “AT08” algorithm, in making personalized commodity rankings, by comparing their ranking results with respect to the original eBay search result ranking. The comparison results are reported in Table 3, which are visually compared in Figure 4. These comparison results confirm our above hypothesis that our personalized webpage ranking method works more effectively for domain-specific searches.

7. DISCUSSION AND FUTURE WORK

In this paper, we introduce a new webpage recommendation algorithm, which utilizes users’ facial expressions and

gaze positions during their webpage browsing activities as implicit user feedbacks for recommending interesting webpages to a user in a personalized way. Given a set of facial expressions and gaze position samples of a user captured during his or her webpage browsing activities, we first measure the user’s interest on individual words in the webpages that he or she has previously browsed. Based on the estimated word-level user interest, our algorithm then predicts the potential interest of the user for all the webpages the user has not previously browsed in a web query’s search result set. By ordering these webpages according to the predicted user interest, our algorithm can make a personalized webpage recommendation to suit for the user’s interest. To evaluate the effectiveness of our algorithm, we compare the personalized webpage ranking produced by our algorithm with the ideal webpage ranking expected by the user. We also examined the rankings of webpages produced by a few peer algorithms. Through this set of experimental studies, we demonstrated that our personalized webpage recommendation algorithm can yield a webpage ranking that better reflects a user’s expected ideal webpage rankings than all the peer algorithms.

As previously discussed, the “reading interest” of a user plays an crucial role in our algorithm. In our current method, we do not differentiate between webpages of different contents when capturing a user’s reading concentration data. We expect however that a user’s reading behavior over say novels versus a weather report would be very different, due to different reading motivations and purposes. A more careful and sophisticated treatment of these training data for better prediction accuracy will be within the scope of our immediate future work.

A natural extension to our work is to apply the main ideas of our algorithm to personalized recommendation for im-

ages and videos. For web images or videos with annotation texts, we can anchor them onto their associated texts and then apply our algorithm to the annotation texts. For those images or videos without text tags, we may first compute the similarity between images or videos based on their visual features, e.g., following the method suggested in [16]; and then we can monitor a user's reading concentration during his or her previous image browsing and video watching activities to acquire training data for predicting the user's potential reading interest on a new image or video by adopting the same type of algorithmic framework proposed in this paper.

We also plan to extend our algorithm for processing non-English webpages. So far, our algorithm relies on the lexical similarity based on WordNet to propagate word-level user interest to estimate the potential interest of a user on the words or webpages he or she has not previously browsed. Such a choice of the underlying algorithm restricts our algorithm to only be able to deal with English webpages. In the future, we plan to design an adaptive algorithm which can deal with multi-lingual users as there are at least four billion people who are bilingual or multilingual according to Yahoo! answers. Analyzing user web browsing behaviors over multilingual materials can provide valuable semantic clues on what may be the true interest of a user. Such extension work can not only make our algorithm more generically applicable for recommending webpages in multiple languages, but can also help improve the recommendation accuracy for webpages in a single language.

Furthermore, we have noticed that the ontology provided by WordNet only represents the lexical relations between words, but not their in-depth semantic relationships. We observe that in our current experiments, those keywords with most significant SES reduction rates achieved by our algorithm usually have specific meanings. This indicates that our algorithm can achieve a better performance for executing queries with clear search intents, but not as effective when the query intent is vague or ambiguous. This phenomenon is partially due in the fact that it is difficult for our algorithm to discover interesting contents on its own for topics to which the user has not shown interest previously. However, when a query word is ambiguous or refers to a large scope of potential meanings, our algorithm may inadvertently use training data from other topics which are not closely related to the user's current search intent to predict the user's search interest. In those circumstances, our algorithm would produce noisy recommendations. We suspect that using a more powerful ontology for analyzing readers' interest following a semantics and ontology based approach rather than our current word-based approach could allow us to more accurately estimate user interest distribution for making more precise personalized webpage recommendation. In the future, we plan to look into Wikipedia in this regard [7].

Finally, in comparison with making personalization for general web searches, personalized recommendation for domain-specific searches may be more effective and useful, especially for many e-commerce applications such as online advertisements or targeted product information distribution. In this paper, we have demonstrated the superiority of our method for domain-specific search recommendation through compar-

ing with eBay, which has yielded very encouraging results. In the future, we intend to explore more tailored knowledge representation and reasoning methods to enhance our algorithm's capability for generating personalized product information recommendation.

Acknowledgement

This work has a patent pending.

8. REFERENCES

- [1] eMotion, Visual Recognition. <http://www.visual-recognition.nl>, 2006-2008.
- [2] A. Bulling, J. A. Ward, H. Gellersen, and G. Tröster. Robust recognition of reading activity in transit using wearable electrooculography. In *Pervasive '08: Proceedings of the 6th International Conference on Pervasive Computing*, pages 19–37, 2008.
- [3] A. Bulling, D. Roggen, and G. Tröster. It's in your eyes: towards context-awareness and mobile HCI using wearable EOG goggles. In *UbiComp '08: Proceedings of the 10th International Conference on Ubiquitous Computing*, pages 84–93, New York, NY, USA, 2008. ACM.
- [4] E. H. Chi, M. Gumbrecht, and L. Hong. Visual foraging of highlighted text: An eye-tracking study. In *HCI '07: Proceedings of HCI International Conference*, pages 589–598, 2007.
- [5] P. A. Chirita, C. S. Firan, and W. Nejdl. Personalized query expansion for the web. In *SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 7–14, New York, NY, USA, 2007. ACM.
- [6] Z. Dou, R. Song, and J.-R. Wen. A large-scale evaluation and analysis of personalized search strategies. In *WWW '07: Proceedings of the 16th International Conference on World Wide Web*, pages 581–590. ACM, 2007.
- [7] E. Gabrilovich and S. Markovitch. Computing semantic relatedness using wikipedia-based explicit semantic analysis. In *IJCAI'07: Proceedings of the 20th International Joint Conference on Artificial Intelligence*, pages 1606–1611, San Francisco, CA, USA, 2007. Morgan Kaufmann Publishers Inc.
- [8] D. Gorodnichy. Perceptual cursor—a solution to the broken loop problem in vision-based hands-free computer control devices. *National Research Council Canada Publication*, NRC-48472:1–23, 2006.
- [9] T. H. Haveliwala. Topic-sensitive pagerank. In *WWW '02: Proceedings of the 11th International Conference on World Wide Web*, pages 517–526, New York, NY, USA, 2002. ACM.
- [10] B. J. Jansen, A. Spink, J. Bateman, and T. Saracevic. Real life information retrieval: a study of user queries on the web. *SIGIR Forum*, 32(1):5–17, 1998.
- [11] K. Järvelin and J. Kekäläinen. Ir evaluation methods for retrieving highly relevant documents. In *SIGIR '00: Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 41–48, New York, NY, USA, 2000. ACM.

#	Search Keywords	eBay	Google		AT08		Ours	
		SES	SES	Reduction(%)	SES	Reduction(%)	SES	Reduction(%)
1	sweet juice	234	198	15%	174	26%	151	35%
2	megatron	96	96	0%	75	22%	70	27%
3	wow model	346	381	-10%	201	42%	164	53%
4	video card game	165	175	-6%	143	13%	127	23%
5	puerto rico	253	264	-4%	163	36%	148	42%
6	texas card	62	68	-10%	52	16%	50	19%
7	one two infinity	74	87	-18%	64	14%	52	30%
8	wild china	118	125	-6%	92	22%	88	25%
9	black jacket	127	123	3%	105	17%	101	20%
10	crane	142	150	-6%	118	17%	92	35%
Average		162	167	-4%	119	22%	104	31%

Table 3: Ten example searches’ corresponding search error scores (SESs) for the rankings produced by eBay (by “Best match”), Google, the “AT08” algorithm, and our algorithm respectively. For the latter three methods, we also calculate their respective SES reduction ratios with respect to eBay’s ranking. Both the “AT08” algorithm and our algorithm have access to the user data collected from the volunteers in the first week of our experiments as the training data. Key statistics of the SES reduction ratios are also visually compared in Figure 4 using boxplots.

- [12] K. Järvelin and J. Kekäläinen. Cumulated gain-based evaluation of ir techniques. *ACM Transactions on Information Systems*, 20(4):422–446, 2002.
- [13] K.Thomas. Amygdala response to facial expressions in children and adults. *Biological Psychiatry*, 49(4):309–316, 2001.
- [14] Y.-P. Lin, Y.-P. Chao, C.-C. Lin, and J.-H. Chen. Webcam mouse using face and eye tracking in various illumination environments. *EMBS ’05: Proceedings of 27th IEEE Annual International Conference of Engineering in Medicine and Biology Society*, pages 3738–3741, 2005.
- [15] C. M. Loba. Enable Viacam, CREA Software Systems. <http://eviacam.sourceforge.net>, 2008-2009.
- [16] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [17] G. A. Miller. WordNet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.
- [18] D. K. Ming-Hsuan Yang and N. Ahuja. Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 24(1):34–58, 2002.
- [19] A. Ntoulas, M. Najork, M. Manasse, and D. Fetterly. Detecting spam web pages through content analysis. In *WWW ’06: Proceedings of the 15th International Conference on World Wide Web*, pages 83–92, New York, NY, USA, 2006. ACM.
- [20] F. Qiu and J. Cho. Automatic identification of user interest for personalized search. In *WWW ’06: Proceedings of the 15th International Conference on World Wide Web*, pages 727–736, New York, NY, USA, 2006. ACM.
- [21] R. Ruddaraju, A. Haro, K. Nagel, Q. T. Tran, I. A. Essa, G. Abowd, and E. D. Mynatt. Perceptual user interfaces using vision-based eye tracking. In *ICMI ’03: Proceedings of 5th International Conference on Multimodal Interfaces*, pages 227–233, New York, NY, USA, 2003. ACM.
- [22] A. Sieg, B. Mobasher, and R. Burke. Web search personalization with ontological user profiles. In *CIKM ’07: Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management*, pages 525–534. ACM, 2007.
- [23] M.-C. Su, S.-Y. Su, and G.-D. Chen. A low-cost vision-based human-computer interface for people with severe disabilities. *Biomedical Engineering Applications, Basis, and Communications*, 17:284–292, 2005.
- [24] J.-T. Sun, H.-J. Zeng, H. Liu, Y. Lu, and Z. Chen. Cubesvd: a novel approach to personalized web search. In *WWW ’05: Proceedings of the 14th International Conference on World Wide Web*, pages 382–390, New York, NY, USA, 2005. ACM.
- [25] P. Vuilleumier and S. Schwartz. Emotional facial expressions capture attention. *Neurology*, 56:153–158, 2001.
- [26] I. H. Witten, A. Moffat, and T. C. Bell. *Managing gigabytes: compressing and indexing documents and images*. Morgan Kaufmann, 1999.
- [27] S. Xu, H. Jiang, and F. C. Lau. Personalized online document, image and video recommendation via commodity eye-tracking. In *RecSys ’08: Proceedings of the 2nd ACM Conference on Recommender Systems*, pages 83–90, 2008.
- [28] S. Xu, H. Jiang, and F. C. Lau. User-oriented document summarization through vision-based eye-tracking. In *IUI ’09: Proceedings of the 13th International Conference on Intelligent User Interfaces*, pages 7–16, New York, NY, USA, 2009. ACM.
- [29] S. Xu, Y. Zhu, H. Jiang, and F. C. M. Lau. A user-oriented webpage ranking algorithm based on user attention time. In *AAAI ’08: Proceedings of the 23rd AAAI Conference on Artificial Intelligence*, pages 1255–1260, 2008.
- [30] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, 2003.
- [31] Q.-F. Zheng and W. Gao. Constructing visual phrases

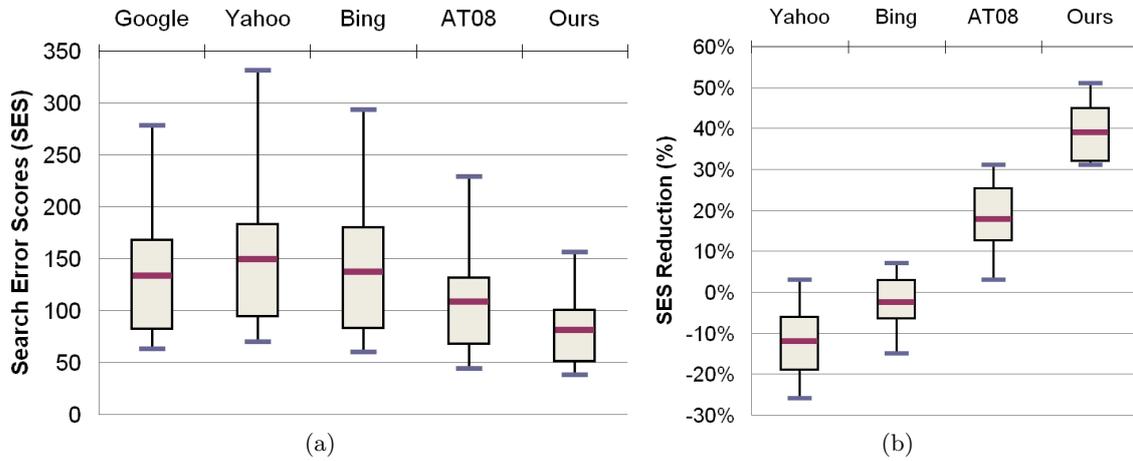


Figure 3: (a) Boxplots on search error scores of webpage rankings produced by the “AT08” algorithm, our algorithm, and three popular commercial web search engines—Google, Yahoo!, and Bing respectively. (b) SES reduction ratios for the rankings produced by Yahoo!, Bing, the “AT08” algorithm, and our algorithm respectively with respect to Google’s webpage ranking. Meanings of key elements in our boxplot have been previously illustrated in Figure 2(d).

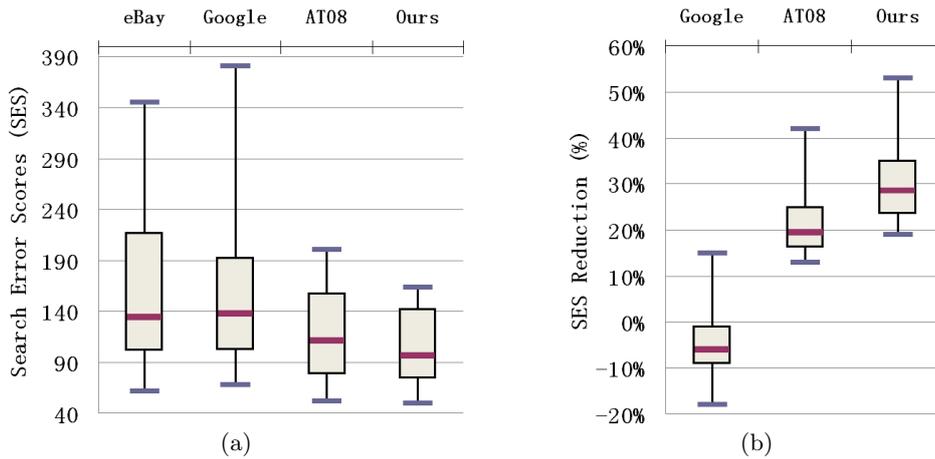


Figure 4: (a) Boxplots on search error scores of commodity rankings produced by eBay, the “AT08” algorithm, our algorithm, and Google respectively. (b) SES reduction ratios for the rankings produced by Google, the “AT08” algorithm, and our algorithm respectively, with respect to eBay’s original commodity ranking (by “Best match”). Meanings of key elements in our boxplot have been previously illustrated in Figure 2(d).

for effective and efficient object-based image retrieval.
*ACM Transactions on Multimedia Computing
 Communications and Applications*, 5(1):1–19, 2008.

- [32] P. Zielinski. Opengazer: open-source gaze tracker for ordinary webcams, Samsung and The Gatsby Charitable Foundation.
<http://www.inference.phy.cam.ac.uk/opengazer/>, last visited: December 2007.