

## Automated Web Site Evaluation – An Approach Based on Ranking SVM

Peng Li  
Tokyo Institute of Technology  
liorlee@nii.ac.jp

Seiji Yamada  
National Institute of Informatics  
seiji@nii.ac.jp

### Abstract

*This paper proposes an automated web site evaluation approach using machine learning to cope with ranking problems. Evaluating web sites is a significant task for web service because evaluated web sites provide useful information for users to estimate sites' validation and popularity. Although many practical approaches have been taken to present a measuring stick for web sites, their evaluation functions are set up manually. Thus, we develop a method to obtain evaluation function using Ranking SVM and automatically rank web sites with the learned classifier. Also we conducted experiments and confirmed the effectiveness of our approach and its potential in performing high quality web site evaluation.*

### 1. Introduction

The estimated Internet user is over 1.5 billion and the number of web sites has been growing at an exponential rate. In the April 2009 survey, there are about 231 million web sites available to users. Web sites are rich information sources with high update rates. It's extremely easy, cheap and fast to publish on the Internet. But due to the fundamental nature of the web, questionable information is being used recklessly, without adequately assessing its authority and good information is being disregarded, because trust in the information is lacking. This leaves a large responsibility on the user to carefully and critically evaluate the web sites.

As is widely alleged, there are no unified criteria that can be fully trusted for web site evaluation. Each user has different criteria, and it's almost impossible to get a positive result using the same evaluation criteria to the web sites belong to different categories. Moreover, even for the same web site, the evaluation changes by time. Although many tools in evaluating

usability, accessibility and security of web sites start to appear, the evaluation tasks are basically carried out by hand. There are limitations to evaluate enormous amount of web sites manually, so an automated evaluation system is significantly necessary.

In this paper, we propose a method to evaluate web sites rely on categories automatically. Specifically, we take the following 2 steps. First, we focus on the web sites that are classified into categories and establish an evaluation criterion for each category by applying machine learning technique dealing with ranking problems. Then, collect the features of the evaluation objectives and do the evaluation based on these features and the criterion of the corresponding category.

To establish the evaluation criteria, it's important to manage to get the evaluation data from the general public automatically on a regular basis. Until recently, the evaluation tasks are carried out by questionnaire survey. However, the situation has changed with the availability of the registration ranking sites. A registration ranking site is a real-time web site ranking service based on users' recommendations and votes. It has developed significantly during this year. The most popular ones like *SiteRank*<sup>TM</sup> have hundreds of thousands of web sites registered, and users are making categories, registering new sites, voting and writing comments frequently. In this research, we use these categorized ranking information of registration ranking sites to establish the evaluation criteria of web sites. In other words, we treat automated web site evaluation as a classification learning problem which learns the discriminant functions from the categorized ranking information as training data. Since we need an appropriate learning algorithm to solve this problem, we use Ranking SVM [1] as our learning algorithm due to its high accuracy of learning totally ordered multi-class discriminant functions.

## 2. Related Work

Web site evaluation demands appropriate evaluation measures. Researches such as Web WISDOM [2] and WebMac [3] use the index achieved from the findings of library and information science as meta-information to make the evaluation measures. But these researches are not aimed as establishing evaluation criteria automatically.

Meanwhile, aside from the approach of questionnaire survey, Velayathan and Yamada [4] proposed an approach to automatically evaluate web pages based on user browsing behavior. However, the evaluation objectives in this research are web pages, not web sites, and the discriminant functions achieved can only handle 2 classes such as “interested” and “not interested” which are not appropriate for web site evaluation.

Moreover, many tools focus on the evaluation of usability, accessibility and security of web sites become available recently. These tools can be divided into 5 categories: 1. Analysis of server performance; 2. Analysis of usage(based on log data); 3. Check of guideline compliance; 4. Analysis of navigation text; 5. Navigation simulation by cyber agent. Unfortunately, tools in the comprehensive evaluation of web sites have not been found.

## 3. Automated Web Site Evaluation

### 3.1. Ranking SVM

Fig. 1 shows the graphical representation of the ranking mechanism. The distance from the hyperplane  $w$  of a datapoint  $x$  is mapped to a one dimensional space. In this space,  $\theta_1, \dots, \theta_{K-1}$  are the different thresholds against which the distance is compared. Data point  $x$  having rank  $i$  will satisfy  $\theta_{i-1} < w^T x_j < \theta_i$ . On the contrary, checking which section the value of  $w^T x_j$  go into will help determine the rank of  $x$ .

### 3.2. Evaluation Features

In this research, we use the ranking information

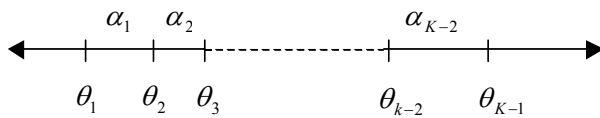


Fig. 1. Ranking mechanism.

Table 1. Evaluation features classified in 10 fields.

Field	Number of features
Top page's global link popularity	1
Freshness	2
Indexable text information	1
Multi-media contents	4
Accuracy of spelling and grammar	1
Accuracy of HTML documents	4
Contents security	5
Contents constitution	4
Design	5
Others	4

from registration ranking sites as training data. Each data is composed of multiple evaluation features and their values, and the data value is the rank determined by the ranking sites. Meanwhile, using the discriminant functions learned from the training data, we can rank the unranked web sites according to their features. Data of these unranked web sites are test data.

The traditional evaluation method of questionnaire survey can deal with both user's subjective and objective features. But in this research, our goal is to automate the evaluation process. Thus the features we need have to be objective and can be obtained automatically from web servers. Table. 1 shows the list of such 31 features categorized in 10 fields which are used in our research at the current stage. Detailed information on each feature is described below.

**Global Link Popularity.** Global link popularity shows the popularity of a web page by examining the link structure of the web. We use Google PageRank here.

**Freshness.** The freshness of a web site is the contrast of new information and old information. We use update frequency and changed amount of text information as a guide.

**Indexable Text Information.** It's the amount of visible HTML text information obtained by calculating the size of HTML source file with tag information removed.

**Multi-media Contents.** The number of files including images, videos, audios and flashes are used here.

**Accuracy of Spelling and Grammar.** The number of misspellings is counted at the current stage.

**Accuracy of HTML Documents.** Several features such as whether the character code and the size of images are specified or not are checked.

**Contents Security.** Features related to the control of cache, scripts and web robots are checked here.

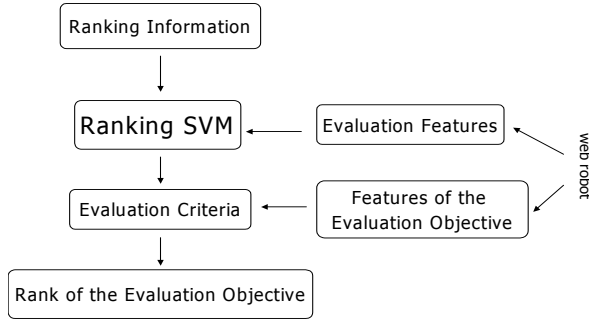


Fig. 2. System construction.

**Contents Constitution.** Proportions of multi-media contents to text information are used as a guide.

**Design.** Information about BGM, background image, frames, style sheets, etc are considered here.

**Others.** Features such as the number of links on the top page, the number of links overall, whether the author of the page is specified and whether the description of the page is specified are used here.

### 3.3. Web Site Evaluation Based on Ranking SVM

Fig.2 shows the structure of our web site evaluation system based on Ranking SVM. At the learning phase, categorized ranking information obtained from registration ranking sites and evaluation features collected by web robots are input to Ranking SVM to learn categorized discriminant functions. At the test phase, unranked web sites get predicted ranks automatically according to their features and the corresponding discriminant function.

## 4. Experiments

We conducted experiments to verify the effectiveness of the automated web site evaluation based on Ranking SVM. The data we used were from a total of 735 web sites classified into seven categories. As discussed before, 31 evaluation features are selected. We use libSVM2.84 as a SVM tool.

In addition, we compare the performance of Perceptron-based ranking algorithm (Pranking [5]), multi-class SVM and Ranking SVM to verify whether Ranking SVM is adapted to our research.

### 4.1. Evaluation Measures

We evaluate web sites with a 5-level evaluation: 1 (Excellent) to 5 (Not expected). We use the mean absolute error (MAE) to evaluate the system

performance. In statistics, the mean absolute error is a quantity used to measure how close forecasts or predictions are to the eventual outcomes.

### 4.2. Data Preprocessing

The values of evaluation features vary from 0 to millions, so before the experiment, we scaled the data to [0,1]. Then we extended data vector to  $n + K - 2$  dimensions to apply Ranking SVM. Here  $K$  is 5, so we did a 3-dimension extension. Most of the ranking information is obtained in the form of totally ordered rank. We transformed the ranking information from total order into partial order by dividing data into 5 equally sized groups and labels are assigned to them ranging from [1,5].

LibSVM provides four kinds of kernels: linear, polynomial, radial basis function and sigmoid. We did a preliminary experiment to investigate the accuracy of all these kernels. As a result, we decided to use the RBF kernel in the future experiment due to its highest accuracy. RBF kernel has two parameters,  $c$  and  $\gamma$ . We did a grid search using cross-validation to find the best  $c$  and  $\gamma$  for each of the seven categories.

### 4.3. Experimental Result

Fig.3 shows the experimental result of our evaluation experiment. 10-fold Cross-validation was conducted for each category in order to avoid influence of biased data. The average of MAE is 0.78, and the standard deviation is 0.3. That is, in most of the cases, the predicted value is either equal to the true

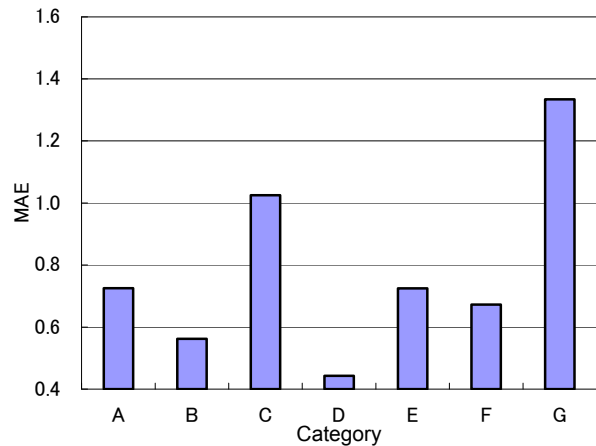


Fig. 3. Result of evaluation experiment.

The y-axis represent the values of MAE, and the x-axis correspond to the categories A: computer science, B: music, C: health and diet, D: movies, E: sports, F: games, G: fashion.

Table 2. Result of comparative experiment.

	MAE	Minimum number of adjacent transpositions
Ranking SVM	0.78	132.7
Multi-class SVM	0.93	217.4
Pranking	1.12	264.5

value or has a shift of 1 rank. We investigated several studies which also use MAE with 5-level evaluation and found that it's considered to be sufficiently effective if MAE is held under 1.

In addition to the perceptron-based ranking algorithm, ranking can be treated as a multi-class classification problem. Table.2 shows the result of comparative experiment of the three algorithms. Note that to get more detailed information, besides MAE, the performance of the three algorithms was analyzed by computing the minimum number of adjacent transpositions needed to bring the predicted value to the ground truth. It is clear from Table.2 that Ranking SVM significantly outperformed perceptron-based ranking algorithm and multi-class SVM. Hence it means that Ranking SVM is more adapted to our research than other ranking algorithms. It also implied that the ranking problem in this research is a nonlinear problem.

## 5. Discussion

From the experimental results, we can see that the gaps of the evaluation accuracy between categories are not small. The reason why categories with low accuracy exist can be considered from two aspects. First, it's difficult to set unified evaluation criteria for some categories because of their characteristics. For example, for health and diet category, the purpose of the users visiting the web sites may be very different from that of users for movie and music categories. They may be looking for health advice, foods, recipes or even body building equipments. Naturally, different purposes lead to different evaluation criteria. Second, it's possible that one or more important features which are key to the establishment of evaluation criteria are not acquired in the experiment. For instance, if beautiful images with high resolution are preferred in fashion category, then it's difficult to set a high accuracy criterion without feature describes it.

Data used in this research are from over 700 web sites. In order to get enough data for each category, we collected data from multiple registration ranking sites. This may lead to the decrease of evaluation accuracy

due to the change of user group and the biased quality of web sites. With the development of registration ranking sites, enough data will become available from a specific site, and the improvement of evaluation accuracy can be expected.

In addition, in the registration ranking sites, categories tend to be more and more segmentalized. It's believed that the evaluation accuracy will grow if the categories become more specialized. Furthermore, as mentioned before, with the improvement of web robot and the development of web technology, more evaluation features will become available. With more adapted features acquired, things will change for the low-accuracy categories.

## 6. Conclusions

In this paper, we proposed an automated approach to evaluate web sites. Ranking SVM was used to learn discriminant functions through categorized ranking information obtained from registration ranking sites and evaluation features collected by web robots. Unranked web sites can be evaluated automatically according to their evaluation features and the corresponding discriminant function. The experimental results confirmed the effectiveness of our approach and verified that Ranking SVM is an adapted ranking algorithm to our research.

This paper shows a practical approach to give an automatic and comprehensive evaluation to any web site on the web. It also indicates the potential of this approach in performing high quality web site evaluation.

## 7. References

- [1] S. Rajaram, A. Garg, X.S. Zhou, and T.S. Huang, "Classification approach towards ranking and sorting problems", Proc. of ECML, Cavtat-Dubrovnik, Croatia, 2003, pp. 301-312.
- [2] J.E. Alexander, and M.A. Tate, "Web WISDOM: How to evaluate and create information quality on the Web", Lawrence Erlbaum, Hillsdale, NJ, 1999.
- [3] M.P. Arnone, and R.V. Small, "WWW motivation mining finding treasures for teaching evaluation skills grade 1-6", Linworth Publishing, Worthington, OH, 1999.
- [4] G. Velayathan, and S. Yamada, "Behavior-based Web page evaluation", Journal of Web Engineering, 2007, pp. 222-243.
- [5] K. Cramer, and Y. Singer, "Pranking with ranking", Proc of NIPS, MIT Press, 2002, vol. 1, pp. 641-647.