An Intelligent Method to Assess Webpage Quality using Extreme Learning Machine

B. Jayanthi[1]
[1]Research Scholar, Department of Computer Science, Rathnavel Subramaniam College of Arts & Science, Sulur, Coimbatore, India.

P. Krishnakumari[2]
[2]Director, Department of Computer Applications (MCA) Rathnavel Subramaniam College of Arts & Science, Sulur, Coimbatore, India.

Abstract

The rapid growth of the Internet as an environment for information exchange and the lack of enforceable standards regarding the information it contains numerous information quality problems. Website quality is dependent on the quality of the software. In the early years, quality of software provided effective support to develop the websites' performance. However, in the website application of new discipline quality is the major challenging process. There is an increase in the web technology, so there is a need for factors access which is associated with success of website increases as well. Many of the existing website evaluation methods and criteria for evaluating website quality are unable to sufficiently assess the performance and quality of a website, and most of them focus on usability and accessibility. This work proposed an intelligent algorithm based on Extreme Learning Machine (ELM) for evaluating the website quality with respect to the service type which it offers. It is able to evaluate a website which has "best" or "worst" quality by type or by pasting a URL into the text box.

Keywords

Web Page Quality, Support Vector Machine, Extreme Learning Machine, Intelligent Webpage Quality Assessment

1. Introduction

"Websites are vital part of our daily life which is used to convey and exchange information among the user groups. This exchanged information will be of different forms, languages and types and integrates the images, text, video and sound which are intended to sell, inform, and persuade a change in belief or attitude or a viewpoint. But the quality assessment is still challenging research area despite website increase. While user interfacing a website, the quality should meet the customer satisfaction and meet the expectation of accomplishment level of user. This website quality is an important topic in the current research mainly in website quality metrics.

From 1990's, metrics set has been proposed for evaluating the website quality attributes. Although quality of website has valuable background and been well developed in recent years, a big question is "why is the quality of websites still poor and lack of quality features cause user dissatisfaction in most websites". The new website quality features regulate to establish a new website quality metrics

which will have more practical measurement criteria and appropriate approaches for website quality evaluation needs. The aim of this work is to investigate the definition of website quality criteria, look for an approach to do the website evaluation that can relate to the user perspective. Especially developing the practical website quality metrics and identifies each quality characteristic, sub-characteristic and measurable measures. The Extreme Learning Method (ELM) is used for intelligent classification purpose for webpage quality assessment.

Recently, Extreme Learning Machine (ELM) has fascinated the attention of many researchers in different applications. ELM is an advancement of single layer feed forward neural network which is an improved version of standard feed forward neural network. In this work improved MLP (Multi-Layer Perceptron) learning is used which is more efficient training algorithm. The advantage of ELM is obvious in shorter training time and in compact model size (i.e., computer memory to store the trained model) while the generalization of ELM is comparable to that of SVM. In this work, the performances of ELM (with or without prior duplication) in different aspects were evaluated by comparing the results with SVM.

The subsequent sections are organized as follows: related works are given in Section 2, followed by the proposed methodology in Section 3. The experiment details are presented in Section 4. The conclusion is discussed in Section 5 and followed by the references are listed.

2. Literature Survey

Lin (2010) developed a model that integrates fuzzy logic and analytic hierarchy process to develop a hybrid fuzzy model which prioritized the relative weights of course website quality factors. The results indicated that there are some flaws between high and low-experience groups with regard to the evaluation of website quality. The video quality subjective study is presented by Seshadrinathan et al (2010) based on the collection of videos with variety of application relevant processes. A model is proposed by Wells et al (2011) for website quality investigation and considered the effects of product information asymmetric

and signal credibility. Udo et al (2010) reported the website quality dimensions based on his internet experience as e-customer of web service, expectations and perceptions. They developed a quality constructs of web service and their relationships are analyzed based on behavioral intentions and customer satisfaction in an e-business environment.

O'Cass studies the role of consumer perceptions of innovativeness in websites as a factor influencing their evaluations of website service quality, development of trust in the website, loyalty to the website and word-of-mouth behaviors toward the website. The parameters under investigation in the research study were measured using established parameters from the e-business marketing literature and adapted to the context of the research study. The website quality value is measured by Gregg and Steven (2010) based on impact on price premiums, trust and intention to transact. They examine the extent for improving the online auction quality to which the trust can be induced. The interplay among the consumer's inherent is examined by Wells et al (2011) and reports on two empirical studies. An e-banking website quality assessment methodology is proposed by Kahraman and Tolga (2011) based on approach of integrated fuzzy AHP-ELECTRE. To assess the website quality levels the fuzzy ELECTRE is used. Ho et al (2012) contributes empirical data to the primary theoretical literature on web searching behavior in general and web attitudes and website quality in particular. It is, to a certain extent, common sense that website quality has a positive impact on browsing behavior.

Huang et al (2012) studied about the ELM and show the following: 1) ELM can be applied directly in regression and multiclass classification and provides a unified learning platform, 2) by optimization method point of view, this ELM when compared to LS-SVM and PSVM has milder constraints of optimization, 3) this ELM can classify any disjoint regions and continuous function. Huang et al (2010) examined the classification of ELM based on the standard optimization methods and extended an approach to generalized SLFNs. Extreme learning machine (ELM) as a new learning algorithm was proposed by Liu et al (2010) for single-hidden layer feed forward neural networks. This ELM will minimize the training error for achieving the good performance.

Wang et al (2011) proposed an concept of ELM and the enhancement of ELM algorithm is also called EELM. Before calculation of the output weight, the input weight and bias is to be proper by using the EELM. In network the learning rate and robustness property is extended by the improved technique. Jun et al (2011) discussed about negative fuzzy rules into the classification of image and the fuzzy rule of positive and negative is combined to form fuzzy rule system and the classification of remote sensing image and natural image is removed by applying fuzzy rule system. Cao et al (2012) proposes an improved learning algorithm for classification which is referred to as voting based extreme learning machine. The proposed method integrates the voting method into the extreme learning machine for webpage quality classification applications.

3. Research Methodology

From the range of reference discipline, disparate sources and empirical practices, the knowledge and experience are integrated in this research methodology. The measurable features and indicators are identifying by the objectives. A current representation of the perfect website is comprised by the set of features which is developed for this work. The proposed technique can be used to identify a path for improvement of a website, to compare between the qualities of websites, and to provide a guideline for developers and designers when creating new websites. There are 16 metrics are proposed to measure the quality of web pages and that are listed in table I.

Table I. Attributes For Web Page Quality

Metric attributes	Description
Link Element	
Total Link	Total Link on Page
Text Link	Total Text Link
Text Element	
Word Count	Total words on page
Total Body Words	Number of words in sentence
Total Sentence	Number of sentence in paragraph
Total paragraph	Number of paragraphs in body text
Total cluster count	Number of text cluster on page
Image Element	
Total Image	Total Image on page
Alt Image Count	Number of images with ALT clause
No Alt Image Count	Number of images without ALT clause.
Animation Count	Number animated element
Unique Image Count	Number of unique images
Image map Count	Number of image maps on page
Unsized Image count	Number of image without size definition
Color Element	
Total Color	Total color on page
Reading Complexity	
Reading Complexity	Overall Page Readability

The metrics of link element, text element, image element, color element and reading complexity are listed in table I and each attributes are discussed briefly.

A. Word Count

Words are the main elements on the web pages for the user to understand and word counting is a major component of many of the metrics that are measured. A simple definition was chosen, whereby a word is delimited by any non-alphanumeric character. This has some

drawbacks such as hyphenated words being counted as two words (eg-"coworker", "side-effect"). The quoted words such as "they're" will also show up two words.

B. Text Cluster Count

Text clusters are blocks of text separated from other text by whitespace eg; two-column paper has at least two clusters per page, more if paragraph breaks exist. Then detect possible text clusters by finding blocks of text terminated by any tag that interrupts continuous line output. Two or more <br> tags (one <br> tag is only one line break, and thus there is no whitespace separator).

C. Total Link

Links are essential elements of the navigation design. Several usability studies have been conducted to provide the breadth (i.e., how many links are presented on a page), depth (i.e., how many levels must be traversed to find information), and others aspects of the navigation structure.

D. Total Image

Web designers to minimize the number of text colors. Display text and Body text color measures report the number of unique colors used for body and display text. The measure do not assess if different colors are used for body and display text.

E. Reading Complexity

Gunning Fog Index (GFI) was the only readability measure correlated with Web interfaces. This study applies GFI formula to compute reading complexity for each web page.

$$Reading\ Complexity = \left(\frac{total\_word}{total\_sentence} + \frac{total\_fog\_word}{total\_word} \times 100.0 \right) 0.4$$

F. Total Body Word and Total Sentence

The designers need to optimize reading comprehension, the number of words in sentences is minimized and the total number of sentences in paragraph on each web page. The readability of prose text, a sentence should not contain more than twenty words and a paragraph should not contain more than six sentences, guidelines proposed.

G. ALT, No ALT Image Count and Unsized Image Count

The literature on Web design guidelines was researched for additional features, similar to the metrics compilation process conducted. Several guidelines show that all images should have an alt tag.

H. Unique Image Count and Average Animation Count

Features were added that were easily measured, even if experts had not associated those features with design elements. This is justified based on the reasoning that it is inadvisable to discard potential features merely because no expert has proclaimed their usefulness, especially considering the current consistencies in guidelines.

3.1 Proposed Method based on Machine Learning

Website is a most important for everyday life, which is used to exchange information between different users. There is a different types for sharing information such as text, audio, image, persuade, video etc., Main challenging in research is website proliferation and website quality assessment. Quality relates to customer satisfaction and also with the level of accomplishment of user expectation when interfacing a website.

The perceptual aspects is captured by quality which is involved in human-website interaction, these aspects is focused on cognitive royalty of a site and affective. Such focus on quality maps to customer satisfaction assessment and contributes to the website quality as aggregate composite that brings together formal metrics and perceptual user traits.

3.1.1 Extreme learning machine

In the proposed system, neural network training algorithm called the Extreme Learning Machine (ELM) is used and evaluates it for Community Directory Miner (CDM). Extreme Learning Machine, ELM, is a recently available learning algorithm for single layer feed forward neural network. Compared with classical learning algorithms in neural network, e.g. Back Propagation and SVM, ELM can achieve better performance with much shorter learning time. All the parameters of ELMs can be analytically determined instead of being tuned. In theory, this algorithm tends to provide the good generalization performance at extremely fast learning speed. ELM theory shows that the hidden node parameters can be completely independent from the training data.

ELM training always seeks a global optimized solution and avoids over-fitting, so it has the ability to deal with a large number of features. The feature selection problem arising in the supervised classification task is effectively addressed by evaluating a separating plane by minimizing separation error and the number of problem feature utilized. The extreme learning machine approach is

formulated using various norms to measure the margin of separation.

4. Experimental Results

Experimental result is performed for web page quality assessment by using SVM and ELM. The computation performance of accuracy is calculated below.

A. Metric Computation

This section explains a sample of web pages with widely differing characteristics which was used to validate the implemented measures. The actual value of each measure was computed and then used to determine the accuracy of computed results. For each page and each measurement, the numbers of accurate hits and misses as well as the average were determined. We calculate overall reading complexity values for different page types including small, large and medium. These pages are extracted from home page, level 1page (directly access from home page) and level 2 page (cannot directly access from home page).

Figure 1. The Demonstration Page

The result shows that the level of the overall reading complexity value is not dependent upon the page sizes. Therefore, we present that the overall reading complexity values of these results are consistent with this literature. According to the overall reading complexity measurement, we find out that the value of complexity measurement can be increased over the same page size. Therefore, we suggest that each information centric page should include at least five complete sentences in each page. The average hit and miss accuracies for a measure are on the basis of all sample pages and the overall accuracy is computed over the average hit and miss accuracies.

Table II. Comparison of web page element measures

Metrics	Avg. accuracy of the system (%)	Avg. accuracy of Ivory (%)
Total Link	98	97.9
Text Link	99.2	99.8
Word Count	99.5	99.9
Total Body Words	100	100
Total Sentence	100	99
Total paragraph	98.5	-
Total cluster count	80	75.8
Total Image	97	99
Alt Image Count	99	-
No Alt Image Count	100	-
Animation Count	97	99
Unique Image Count	100	-
Image map Count	100	-
Unsized Image count	100	-
Total Color	96	97.6
Reading Complexity	98	98.5

In table II, compare the average accuracy of our computing metric with Ivory based on the same demonstration page which is shown in figure 1. The average accuracy is obtained by calculating the hit and miss accuracy divided by 2. We obtained more extracted accuracy in the measure of Total Link, Total Text Cluster and Total Sentence. The table 2 shows the average hit and misses accuracy of metrics extraction.

$$MissAccuracy = \frac{TrueNegative}{TrueNegative + FalsePositive}$$

$$HitAccuracy = \frac{TruePositive}{TruePositive + FalseNegative}$$

B. Building Prediction Model

This study built a prediction model to predict good and not good web pages from the source of education, finance, news and health. After computing metrics by using automated, it produces raw metrics dataset and we extract the most correlated attributes using feature selection technique and apply it as the input in the ELM model. The six attributes are chosen here are total word, reading complexity, total link, total image, total color and text cluster as input dataset to train and test in this model. The web pages are classified into good pages and not good pages based on the assessed metrics. The metrics captured in the ELM can be used to predict the quality of website design. ELM can be used as a classifier based on the dataset of Web page metrics. In this system, Matlab are used to implement metric computation tool and to build a prediction model.

C. Prediction Accuracy

This section presents the result of prediction model for the linearly separate able case based on the web page element metrics dataset. In this study, webpage elements dataset is

appropriate with linear ELM model because it is simple and single dataset. A class is treated as good if it contained more than one text cluster count To predict the accuracy of the model, it should be applied on different datasets thus we performed ELM model.

Table III. Accuracy for proposed ELM

Method	Accuracy (%)
Ivory	75%
SVM	78%
ELM	89%

Table III shows the accuracy for SVM and ELM. Thus the proposed method of ELM has high accuracy when compare with SVM.
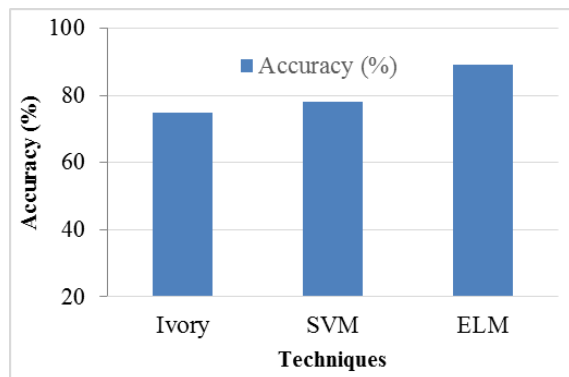

Figure 2. Accuracy for Ivory, SVM and ELM

Figure 2 shows the accuracy for comparison of SVM and ELM. Thus the proposed method of ELM has high accuracy and performs better function for web site quality when compare with SVM.

5. Conclusion

As more and more people have access to the Internet and use Websites for gathering information, it becomes necessary to identify high quality and trustworthy sites. The webpage quality is determined by the measurement criteria to evaluate the attributes for usability and accessibility of a web application. The objective is to make a website useful, accessible, and profitable and user linking. This work reviewed the most recent evaluation technique of ELM, which were used in evaluating the quality of different websites, and proposes a comprehensive framework for assessing the quality of any website regardless of the type of service that it offers. Thus the proposed intelligent method of ELM use list of attributes for improving the quality of the web page assessment.

REFERENCES

[1]. Lin, Hsiu-Fen. "An application of fuzzy AHP for evaluating course website quality." Computers & Education 54, no. 4 (2010): 877-888.

[2]. Seshadrinathan, Kalpana, Rajiv Soundararajan, Alan Conrad Bovik, and Lawrence K. Cormack. "Study of subjective and objective quality assessment of video." Image Processing, IEEE transactions on 19, no. 6 (2010): 1427-1441.

[3]. Wells, John D., Joseph S. Valacich, and Traci J. Hess. "What Signals Are You Sending? How Website Quality Influences Perceptions of Product Quality and Purchase Intentions." MIS quarterly 35, no. 2 (2011): 373-396.

[4]. Udo, Godwin J., Kallol K. Bagchi, and Peeter J. Kirs. "An assessment of customer's e-service quality perception, satisfaction and intention." International Journal of Information Management 30, no. 6 (2010): 481-492.

[5]. O'Cass, Aron, and Jamie Carlson. "An e-retailing assessment of perceived website-service innovativeness: Implications for website quality evaluations, trust, loyalty and word of mouth." Australasian Marketing Journal (AMJ) 20, no. 1 (2012): 28-36.

[6]. Gregg, Dawn G., and Steven Walczak. "The relationship between website quality, trust and price premiums at online auctions." Electronic Commerce Research 10, no. 1 (2010): 1-25.

[7]. Wells, John D., Veena Parboteeah, and Joseph S. Valacich. "Online impulse buying: understanding the interplay between consumer impulsiveness and website quality." Journal of the Association for Information Systems 12, no. 1 (2011): 32-56.

[8]. Kahraman, Cengiz, and Tolga Kaya. "A fuzzy approach to e-banking website quality assessment based on an integrated AHP-ELECTRE method."Technological and Economic Development of Economy 2 (2011): 313-334.

[9]. Ho, Li-An, Tsung-Hsien Kuo, and Binshan Lin. "The mediating effect of website quality on Internet searching behavior." Computers in Human Behavior 28, no. 3 (2012): 840-848.

[10].Huang, Guang-Bin, Hongming Zhou, Xiaojian Ding, and Rui Zhang. "Extreme learning machine for regression and multiclass classification." Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 42, no. 2 (2012): 513-529.

[11].Huang, Guang-Bin, Xiaojian Ding, and Hongming Zhou. "Optimization method based extreme learning machine for classification." Neurocomputing 74, no. 1 (2010): 155-163.

[12].Liu, Nan, and Han Wang. "Ensemble based extreme learning machine." Signal Processing Letters, IEEE 17, no. 8 (2010): 754-757.

[13].Wang, Yuguang, Feilong Cao, and Yubo Yuan. "A study on effectiveness of extreme learning machine." Neurocomputing 74, no. 16 (2011): 2483-2490.

[14].Jun, Wu, Wang Shitong, and Fu-lai Chung. "Positive and negative fuzzy rule system, extreme learning machine and image classification." International Journal of Machine Learning and Cybernetics 2, no. 4 (2011): 261-271.

[15].Cao, Jiuwen, Zhiping Lin, Guang-Bin Huang, and Nan Liu. "Voting based extreme learning machine." Information Sciences 185, no. 1 (2012): 66-77.