Predicting Web Interface Quality Assessment Using Support Vector Machine

Naw Lay Wah
University of Computer Studies, Mandalay
nawnaw09@gmail.com

Abstract

Internet websites are promising as the products for providing services. The key issues related to website engineering are very essential. These websites need to measure and evaluate for quality and for better understanding. Several Metrics were proposed to correspond with items that Web usability guideline associate with good design. In this paper, we investigate empirically the web page quality on the basis of the 16 assessed metrics. We also propose Support Vector Machine (SVM) prediction model to predict the classification of the good web pages and not good web pages. We collect web sites from Webby awards data (2001-2010) and Top Ten PC Magazines. We express the findings of quantitative analysis of web page attributes and how these attributes are calculated. The metrics captured in SVM model can be used to predict the good and the bad of website design.

1. Introduction

There are many design quality guidelines that directly address the quality of static Web pages. The popularity of using the Web translated into an interest by users in creating the maintaining their own Web sites. In an attempt to provide more unambiguous and accessible guidelines, recent research efforts have shifted towards automated approaches for discovering patterns that correspond to Web design quality. Traditional attempt to quantify the quantities of good web site design have used methods of user testing much like that performed for conventional GUIs. When applied to the web, these studies have intended to focus upon design usability. Usability means the ability of a user to successfully use the pages to perform tasks and this term also covers the concept of accessibility. Evaluation of usability is directly related to the success or failure of the user to answer the question.

A key Element of any web site engineering process is metrics. Web metrics are used to better understand the attributes of the web page we create. Metrics are crucial source of information for decision making. A large number of web metrics have been proposed in the last decade to compare the structural quality of a web page. In this paper, we describe simple metrics to measure the web page design elements and display the accuracy of measurement result and compare with the results of other research. We explained why we measure such a kind of metrics from the various aspects of usability guidelines in [15]. In [15], we used linear Support Vector Machine model to classify the web pages into good pages and not good pages on the basics of the 16 assessed metrics. We empirically validate the relationship of web interface metrics and quality of websites using supervised machine learning technique. The computing results are based on webby awards data obtained 2001-2010 and Top Ten PC

Magazines. In this system, we empirically validate web page quality assessment using Support Vector Machine for non-linearly separable case and Weka toolkit is applied. In the next section, we will discuss related works. Section 3 describes the methodology, including the 16 quantitative metrics used and the web page collection. The overview of the system is presented in section 4 and in section 5, experimental result is included. Finally, the paper concludes with discussion and future work.

2. Related Work

Many detailed usability guidelines have been developed for both general user interfaces and for Web page design [20, 16]. Jakob Nielsen's alert box column [4] claims that the top ten mistakes of Web site design include using frames, long pages, non-standard link colors, and overly long download time. Other column provides guidelines on how to write for the web, asserting that since users scan web pages rather than read them, web page design should aid scanability by using headlines, using colored text for emphasis and using 50% less text since it is more difficult to read on the screen than on paper. Although reasonable, guidelines like these are not usually supported with empirical evidence.

Ivory et.al [13] proposed a system that is to search for relationships between the web page metrics and the page's rating. This research follows a similar approach to highly successful work in the area of automated essay grading [9]. Another approach taken by Ivory [14] involved creating a tool named WebTANGO that extracts simple, low-level metrics from web pages such as the number of words, number of links, average image size and so on. This tool is used to extract HTML web pages from Internet and stored in web database. Each web page was classified as good or not good and the researchers used linear discriminant analysis to search for correlations between page metrics and the page's classification. In [15], we proposed 16 metrics to measure the quality of web pages such as HTML, JSP, ASP, PHP and XHTML. We used supervised machine learning technique to find out the relationship between the metrics and web sites quality. The linear SVM prediction model is used to predict the classification of the good and the bad of website design. In [15], we compared the prediction accuracy of SVM model with prior research [14],[17]. In this system, we extend [15] and consider for non-linearly separate able case to predict the web page quality using support vector machine.

Other approaches assess static HTML according to a number of predetermined guidelines, such as whether all graphic contain ALT attributes [1,6]. The Design Advisor uses heuristics about the intentional effects of various elements, such as motion, size images, and color to determine and superimpose a scanning path on a web page [17]. Patrick Peursum [17] also proposed the web interface usability metrics and classified the quality of web pages using machine learning methods. Yogesh Singh et.al [21] proposed the guideline and an automated tool that is developed in JSP and calculated 15 web page metrics based on the Webby Awards.

3. Methodology

3.1 Dependent and Independent Variables

The binary dependent variable in this study is web page quality assessment. The goal of this study is to explore empirically the relationship between web page metrics and web page quality. Web page quality assessment is defined as the probability of good web page design and not good web page design in a class. We use supervised machine learning methods to predict quality of web pages. Our dependent variable will be predicted based on the web page element assessed metrics.

3.2 Empirical Data Collection

In this system, the three criteria are applied to the selection of metrics: (i) metric was strongly related to page classification by [13,14], (ii) metric was considered by web design guidelines authors and (iii) metric was relatively simple to measure. This study calculates quantitative web page metrics (eg., number of words, body text words, number of images, number of links etc) from the web pages that was evaluated for 2001-2010 webby awards [22].

Table 1 lists the metric attributes which were used to measure. These attributes are categorized by collected data source such as Education, Commercial, News and Health. In this table, we added six metrics such as Total Paragraph, ALT Image Count, No ALT Image Count, Imagemap Count, Unique Image Count and Unsized Image Count in order to complete the measure of Ivory. There are altogether 157 attributes which were measured in prior research. As the number of web metrics available in the literature is large, it becomes tedious process to understand the computation of these metrics and draw conclusion and inference from them [17]. A set of 16 metrics is identified and their values are computed for 100 different web sites from Webby Awards Data. The following metrics are considered in this system:

3.2.1 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.

3.2.2 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. According to the literature, text clusters can provide the user's readability and scannability as well as web page quality.

3.2.3 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. Two type of links are image link and text link. Literatures suggest that the total number of links should not exceed more than 20% of overall web page elements.

3.2.4 Total Image

Flanders and Willis [3] encourages 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 [5].

3.2.5 Reading Complexity

The literature survey revealed numerous discussions of the readability or required reading level of text. Spool et al.,[19] determined that the 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.

$$ReadingComplexity = \left(\frac{total\_word}{total\_sentence} + \frac{total\_fog\_word}{total\_word} \times 100.0 \right) 0.4 \quad (1)$$

3.2.6 Total Body Word and Total Sentence

The designers need to optimize reading comprehension, minimize the number of words in sentences and the 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, the guidelines proposed.

Table 1. Proposed Metric

No.	Metrics Attributes	Descriptions
Link Element		
1.	Total Link	Total Link on page
2.	Text Link	Total Text link
Text Element		
3.	Word Count	Total words on page
4.	Total Body Words	Number of words in sentence
5.	Total Sentence	Number of sentence in paragraph
6.	Total Paragraph	Number of paragraph in body text
7.	Text Cluster Count	Number of text cluster on page
Image Element		
8.	Total Image	Total Image on page.
9.	Alt Image count	Number of images with ALT clause.
10.	No Alt Image Count	Number of images without ALT clause.
11.	Animation Count	Number of animated element
12.	Unique Image Count	Number of unique images.
13.	Imagemap Count	Number of Image maps on page.
14.	Unsized Image count	Number of Image without size definition
Color Element		
15.	Total Color	Total Color on page.
Reading Complexity		
16.	Reading Complexity	Overall page Readability

3.2.7 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 porcess conducted by Invory et al. Several guidelines show that all images should have an alt tag .

3.2.8 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.

4. THE SYSTEM ARCHITECTURE

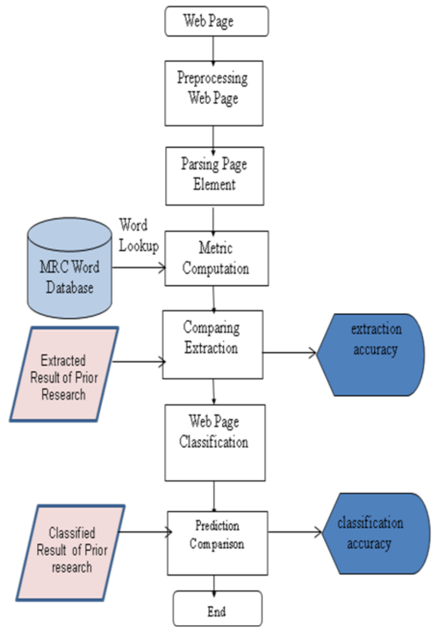

Figure 1. The System Architecture

4.1 Preporcessing , Parsing and Computing Web Metrics

JTidy is the first step of the system to preprocess the web pages. Missing or mismatched end tags are detected and corrected by using JTidy tool. End tags in the wrong order are corrected and fix problems with heading emphasis. It recovers from mixed up tags and adds the missing "/" in end tags for anchors. Prefetch lists by putting in tags missed out and missing quotes around attribute values are added. Moreover, unknown and proprietary attributes are reported. The output of JTidy tool is the input of SAX Parser to make a parsing process. The next step to parse the web interface elements is SAX Parser which means a Simple API for XML and it is Streambased XML parser. It can be used to read and parse XML, HTML and other markup language. It can also parse the individual nodes of the input, reads the input and determines the type of things it encounters. It also recognizes the format of the nested tags. Two types of SAX parser are SAX1.0 and SAX 2.0. The advantages of SAX parser is faster and uses less memory space than DOM parser and HTML parser [7]. In the calculation of metric computing, MRC (Medical Research Council) database [10,12] is used to calculate the Reading complexity measurement.

4.2 Model Prediction Using Support Vector Machines

Support Vector Machine is one kind of learning machine based on statistical learning theory. The basic idea of applying SVM to pattern classification can be stated that it maps the input vectors into one feature space, either linearly or non-linearly, which is relevant with the selection of the kernel function. Then, within the feature space from the first step, seek an optimized linear division, i.e., construct a hyperplane which separates two classes. SVM 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 support vector machine approach is formulated using various norms to measure the margin of separation. In this approach, the most correlated attributes such as total word, reading complexity, total link, total image, total color and text cluster are chosen as input dataset to train and test in SVM model.

4.2.1 The Linear Support Vector Machine

They are a method for creating functions from a set of labeled training data. The function can be a classification function (the output is binary: is the input in a category) or the function can be a general regression function. It is successfully applied in applications such as face identification, medical diagnoses, text classification, pattern recognition and identification of organisms.

In this system, the SVM model is used to classify the quality of web pages based on the web page element metrics dataset. The positive and negative data points are linearly and non linearly separable. The data can be described by ***l*** points in the ***d***- dimensional space. The input $x_i \in R^d$ with two different labels and output $y_i \in \{-1, +1\}$ depending on the class that is assigned to the point x_i for all $i = 1,\ldots,l$. X_i lie on the hyperplane satisfy $w.\ x_i+b=0$, where $w \in R^d$ is normal to the hyperplane, $\frac{|b|}{\|w\|}$, describes the perpendicular distance from the hyperplane to the origin, $\|w\|$ is the Euclidean norm of w. Maximum margin hyperplain is

$$\langle w.x\rangle + b = \sum_{i \in sv} y_i \alpha_i \langle x_i.x\rangle + b = 0 \qquad (2)$$

4.2.2 The Non-Linear Support Vector Machines

In case that SVM cannot linearly separate two classes, SVM extends its applicability to solve this problem by mapping input data into higher dimensional feature spaces using a nonlinear mapping ϕ, such that $x \mapsto \phi(x)$, where $\phi : R^n \to R^m$ is the feature map. It is possible to create a hyperplane that allows linear separation in high-dimensional space. This corresponds to a curved surface in the lower-dimensional space: the transformation from lower to higher dimensional feature spaces using ϕ. This transformation can be done using a kernel function. Therefore, the kernel function is an important parameter in SVM. The Kernel function $K(x_i,x_j)$ is defined as follows:

$$K(x_i, x_j) = \phi(x_i)^T \phi(x_j) \tag{3}$$

Where $K(x_i, x_j)$ is the kernel function performing the non-linear mapping into feature space. Once the solution has been found, the decision can be constructed as:

$$f(x) = sign\sum_{i=1}^{n} \lambda_i y_i k(x_i, x) + b \tag{4}$$

The most common kernel functions are Linear, Polynomial, Gaussian (RBF) and Sigmoid (MLP). The recommended kernel function is the Radial basis Function (RBF) [18]. We used Linear, Polynomial and RBF function in SVM modeling to predict web page quality in this study. The kernel maps non-linearly data into a higher dimensional space, so it can handle non linear relationships between the dependent and the independent variables. Figure 1 shows the sample example of RBF kernel. One category of the other as triangles. The shaded circles and triangles are support vectors. The dependent variable is shown as rectangles and the Given a set of $(xi, yi),\ldots, (xm, ym)$ and $yi \in \{-1,+1\}$ training samples. αi =(i=1,……,m) is a lagrangian multipliers. K(xi,yi) is called a kernel function and b is a bias. The discriminant function D of two classes SVM is given below [8]:

$$x = \begin{cases} +1 & if\ D(x) > 0 \\ -1 & if\ D(x) < 0 \end{cases} \tag{5}$$

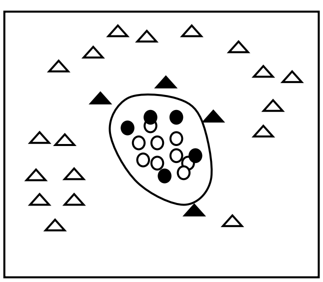

Figure 2. Radial Basic Function

4.3 Prediction Performance Measures

The performance of prediction models for two-class problem (eg: good pages or not good pages) is typically evaluated using a confusion matrix, which is shown in Table . In this study, we used the commonly used prediction performance measures: accuracy, precision, recall and F-measure to evaluate and compare prediction models quantitatively. These measures are derived from the confusion matrix [8].

Table 2. A confusion matrix

		Predicted	
		Good Design	Not-Good Design
Actual	Good Design	TN=True Negative	FP=False Positive
	Not-Good Design	FN=False Negative	TP=True Positive

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

$$F - measure = \frac{2 * Precision * Recall}{Precision + Recall} \tag{9}$$

5. Empirical Evaluation

This section describes the conducted empirical study that evaluates the capability of SVM in predicting web page quality assessment. We used the open source WEKA machine learning toolkit to conduct this study. The parameters for SVM prediction model was initialized mostly with the default settings of the WEKA toolkit. The regularization parameter (C) was set at 1; and the bandwidth (γ) of the kernel function was set at 0.5.

5.1 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. In [15] we calculated and presented overall reading complexity values for different page types including small, large and medium pages.

5.2 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 tool, it produces raw metrics dataset and we extract the most correlated attributes using feature selection technique and apply it as the input to SVM model in WEKA toolkit. The six attributes such as total word, reading complexity, total link, total image, total color and text cluster are chosen 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 Support Vector Machine model can be used to predict the quality of website design. Linear Support Vector Machine can be used as a classifier based on the dataset of Web page metrics.

5.3 Analysis Result

Table 3 shows the accuracy, precision, recall, and F-measure of the model predicted. The model was applied to 145 classes and figure 3 presents the accuracy comparison of the linear, polynomial and Gaussian kernel prediction model. According to the results, Gaussian (RBF) kernel gets the highest prediction accuracy. In figure 3, we present the comparison of the prediction results of applied kernel methods. We investigated that the prediction accuracy of prior approach is 63 percent in [14] and 73 percent in [17] respectively. According to the results of prediction accuracy, an improved approach for web page quality assessment is SVM model. Figure 4 describes the comparison result of this system with prior research.

Table 3. Prediction Result of SVM

Kernel Type	Accuracy	Precisio n	Recall	F-Measur e
RBF	80%	79%	79%	79%
Polynomial	78%	79%	78%	78%
Linear	75%	75%	75%	75%

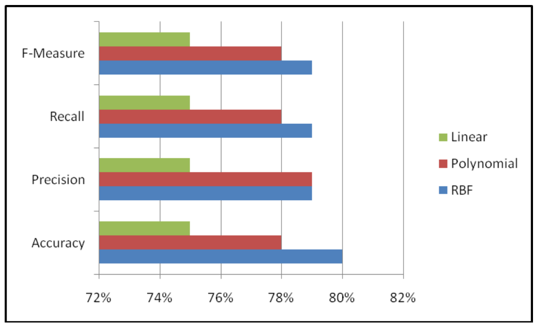

Figure 3. Model Predicted Comparison

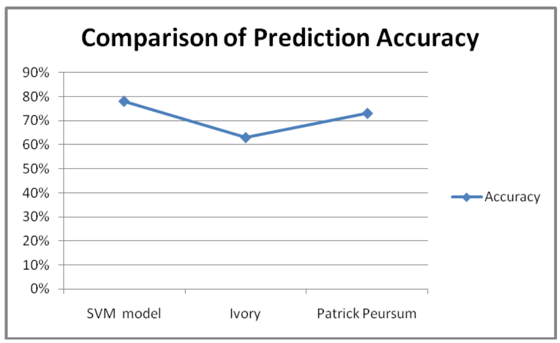

Figure 4. Comparison of Prediction Accuracy

6. Conclusion

This system extends the set of metrics proposed by Ivory to cover additional features. The six metrics are added: Total Paragraph, ALT Image

Count, No ALT Image Count, Imagemap Count, Unique Image Count and Unsized Image Count in order to complete the measure of Ivory. This system described low level measure across the web interface design elements by using automated tool. The goal of this proposed system is to empirically analyze the performance of SVM method based on web page design metrics. The SVM prediction model is used to predict the quality of Web pages. The most correlated attributes such as total word, reading complexity, total link, total image, total color and text cluster are chosen as input dataset to train and test in this model. According to the results by using WEKA machine learning toolkit to conduct this study, the prediction accuracy of SVM model is an efficient model to improve the web page quality assessment. This study will be replicated on the larger dataset and will be calculated the computation complexity and scalability for prediction model.

References

[1] CAST. Bobby, http://www.cast.org/bobby/,2000H.

[2] Cortes, C., Vapnik, V. (1995):" Support-vector networks". *Machine Learning 20, pp.273297.*

[3] Flanders, Vincent, and M. Willis. 1998. "Web Pages That Suck: Learn Good Design by Looking at Bad Design". *San Francisco, CA: SYBEX*.

[4] J. Nielsen (2005). "Weblog usability: The Top Ten Design Mistakes*". Jakob Nielsen's Alertbox.*

[5] J. Nielsen. 2000. "Designing Web Usability: The Practice of Simplicity". *Indianapolis, IN: New Riders Publishing*.

[6] J. Scholtz and S. Laskowski. "Developing usability tools and techniques for designing and testing web sites". *In proceedings of the 4th Conference on Human Factors & the Web, 1998.*

[7] "Java Programming Advanced "- *ICTTI, Union of Myanmar*.

[8] K.O.Elish, M.O. Elish(2007) , "Predicting Defect-Prone Software Modules Using Support Vector Machines" . Science Journal.

[9] K. KukichK. (2000). "The debate on automated essay grading ". *IEEE Intelligent Systems, 15(5), 22-27.*

[10] M. Coltheart. 2001. The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology* 33.497-505

[11] M. Stone, (1974): "Cross-validatory choice and assessment of statistical predictions". *J. Royal Stat. Soc., 36,pp.111-147.*

[12] M. Wilson, 1989. "MRC Psycholinguistic Database: Machine Usable Dictionar, Version 2.00".

[13] M.Y Ivory, R.R. Sinha, and M.A.Hearst. (2000a). "Preliminary findings on quantitative measures for distinguishing highly rated information-centric web pages". *In proceedings of the 6th Annual Conference on Human Factors and The Web,* pages 214-228

[14] M.Y.Ivory,R.R. Sinha, and M.A. Hearst. (2000b). "Empirically validated web page design metrics". *Technical report, University of California, Berkeley.*

[15]N. L. Wah (2011),"An Improved Approach for Web Page Quality Assessment ". *In proceedings of the 9th IEEE Student Conference on Research and Development.*

[16] P. J. Lynch and S. Horton, 1999. "Web Style Guide: Basic Design Principles for Creating Web Sites". *Yale University Press*.

[17] P. Peursum, 2001. "Investigation into the use of Simple Metrics for Predicting Web Page Design Rating".

[18] Sherrod, P. (2003) DTreg Predictive Modeling Software.

[19] Spool, T Scanlon, W. Schroeder, Carolyn Snyder, and Terri DeAngelo. 1999. "Web Site Usability: A Designer's Guide".

[20] T. Comber, 1995. "Building Web pages: An HCI perspective". *Proceeding of the First Australian World Wide Web Conference AusWeb'95, pages 119-124, Ballina.*

[21] Y.Singh, R.Malhotra and P Gupta, 2011. "Empirical Validation of Web Metrics for Improving the Quality of the Web Page".*International Journal of Advanced Computer Science and Applications,* Vol 2, No. 5.

[22] http://www.webbyawards.com