



A novel clustering approach to select optimal usability principles for educational websites

Prafulla Bharat Bafna

Symbiosis Institute of Computer Studies and Research (SICSR),
Symbiosis International University (SIU), Atur Centre, Gokhale Cross Road, Model Colony,
Pune –411 016, Maharashtra, INDIA.

Abstract: Usability is the quality measure of a user's interaction with a product. When this product is website its termed as web usability. It is the quality attribute that assesses how easily user interfaces. User interface contributes to improve the design process. Web usability is an approach to make website easy to use for an end-user, without the requirement that any specialized training be undertaken. Website design is the key issue, because how the end user experiences the design is the key to acceptance. On the Web, usability is a necessary condition for survival. If a website is difficult to use, the users will leave the site and the usability will be zero. There are several usability principles mentioned in the literature but all these principles depend upon the application domain. It means that usability principles for commercial websites may not be applicable to educational websites.

Clustering based technique is proposed which extracts major usability principles amongst several by forward feature selection and experts' knowledge. It reduces time, efforts and resources.

Outcomes of cluster quality is used as decision support for the optimum usability principles and clustering on complete data set for effective decision support.. The data set used is related to top 35 educational websites with respect to its usability count.

Keywords: Usability, scale, optimal, entropy, purity, SMSE, clustering

I. Introduction

With advances in Internet technology, website design had become critical issue .The main objective of website is to achieve goals set by application domain. User can be frustrated and may give up if more efforts are spent by him to accomplish a particular task. Confusing website might affect the productivity.

Usability principles are application specific [1].To get optimal principles for educational website a clustering based approach is proposed in which cluster accuracy is used as an indicator to select the set of features. K-means and hierarchical algorithm are selected to confirm optimal principles. The selection criteria are entropy, purity and Sum of Mean Square Error (SMSE).

Considering usability guidelines, top 35 college websites are analyzed out with respect to their usability count [1]. A questionnaire was designed based on website usability guidelines and distributed to users .Each objective question had 4 options. Each option is associated with a scale (1-4).Question weight is calculated according to the selected option. For each website 35 parameters are considered and distance matrix is formed.

Websites represents rows and scale assigned to each usability principle represents a column

Hierarchical algorithm is applied for different combination of features by considering expert's advice and forward feature selection method [7]. The selection criteria is based upon minimum entropy, maximum purity, minimum SMSE is selected. Finally 18 usability principles were selected from 35. These principles can be considered as mostly focused to design educational website.

II. Background

Usability process involves design with its evaluation and follows basic steps as requirement analysis, conceptual design, prototype, production, launch and maintenance [2]. Home page should be created in such a way that it will create positive first impression of a website. The important issues while designing websites are navigation, graphics, screen based control, page layout, etc.[3]. Benefits of planning usability into the website are Higher satisfaction leads to productivity, Completion with success to acquire brand loyalty, A higher rate of repeat users to progress in competition. The challenging task lies in extracting useful information from a large collection of data either from a data warehouse or from a database. [3,8]Generally collected data contains irrelevant or redundant attributes. Classification and clustering do not give accurate result if there are interdependent attributes. Correct feature selection is a fundamental data preprocessing step in data mining. FeatureMine algorithm contains sequence mining and classification algorithms which efficiently handles very large data sets with thousands of items and millions of records.[4] Edie Rasmussen states Cluster analysis is a technique which assigns items to groups based on a calculation of the degree of association between items and groups. Cluster analysis can be used for hierarchical algorithm. Nested data set is produced in which pairs of

items or clusters are connected successively. However, the hierarchical methods are better information retrieval. The commonly used hierarchical methods, such as single link, complete link, group average link, and Ward's method, have high space and time requirements. In order to cluster the large data sets with high dimensionality there is need to have a better algorithm. Examples are the minimal spanning tree algorithms for the single link method, the Voorhees algorithm for group average link, and the reciprocal nearest neighbor algorithm for Ward's method. Edie listed steps of clustering including Selecting of the attributes on which items are to be clustered, selecting appropriate clustering method, Creating the clusters or cluster hierarchies, interpreting clusters and validating the results etc.[5] They have focused on feature selection algorithms for classification (knowing class label) and clustering (unsupervised feature selection) where data is unlabeled. Feature selection algorithms designed with different evaluation criteria broadly fall into three categories: the filter model, the wrapper model and the hybrid model. The filter model relies on [6]

III. Data Collection

Data is. Collected from 35 different websites having topmost usability count [1] and 35 usability principles are considered for each website. Each principle is converted into an objective question. Scale is assigned to each question [1]. K-means [5] plot was obtained for above matrix. Consistency represents X-axis and feedback message represents Y axis. Hierarchical algorithm [5] was applied on entire data and dendrogram was obtained presented in graph A and B.

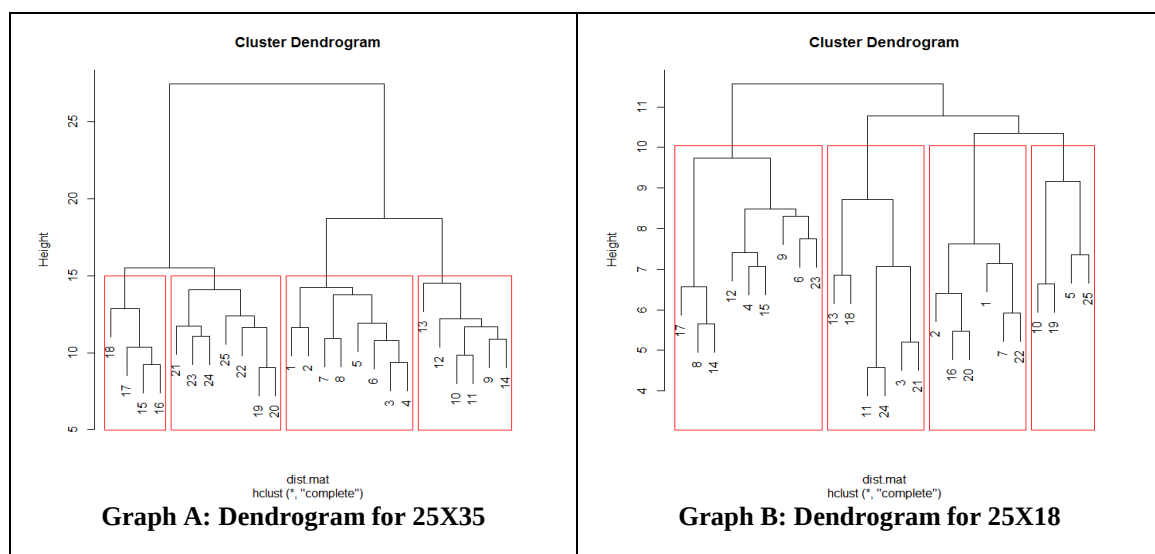
IV. Selection of Optimal usability principles

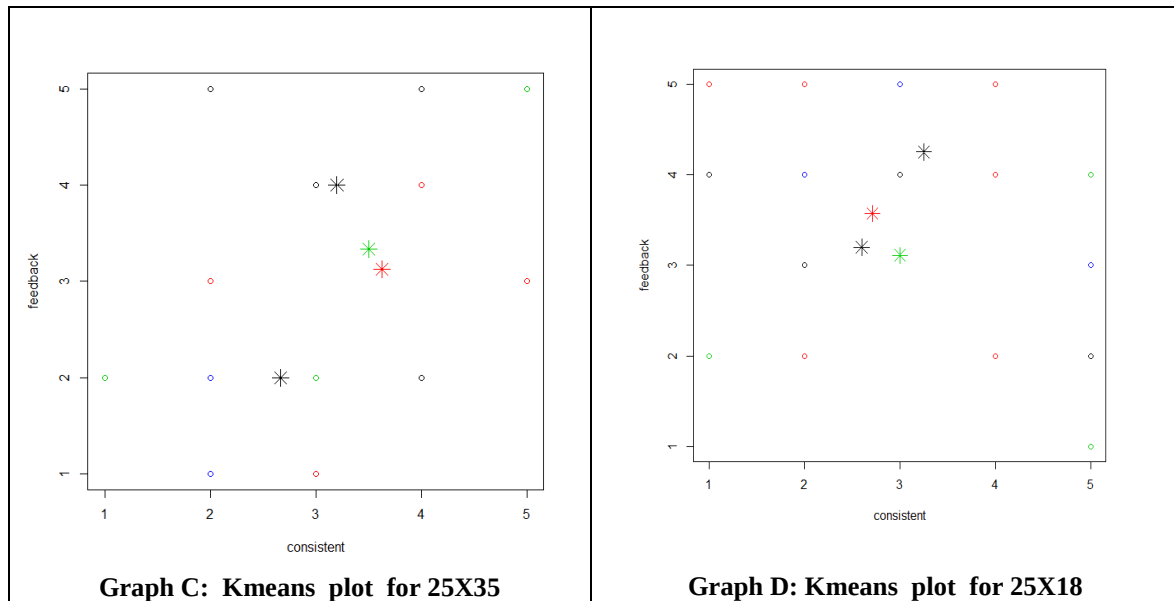
k-means and hierarchical algorithms are applied on domain specific enhanced principles by forward Feature selection. approach.[7]. Graph A,B represents dendrogram produced by hierarchical algorithm using Euclidean distance for 25 websites and 18 usability principles respectively. Graph C,D represents kmeans plot for 25 websites and 18 usability principles respectively.

We can easily detect optimal guidelines form Table A having minimum SMSE and entropy also with maximum purity. From 35 standard website principles, principles selected are are navigation, error free, flexibility, homepage, menu, feedback, memory load, graphs, notice board, category label, validation, scrolling, group label, grouping, response time, unique actions, parallel actions, focus

Table A: Optimal website principles

No of Principles	SMSE (sum mean square error)	entropy	purity
10	15.67	.29	.76
14	20	.35	.78
18	8.26	.19	.88
24	25	.38	.79
30	12	.24	.74
35	>22	.42	.69





V. Conclusion

The overall quality of a website is actually sum of many quality attributes, important one of which is usability. Focusing usability as a quality goal, there are several usability principles explored in literature. These principles are not domain specific. Applying all principles at a time is cumbersome. Above approach of selecting optimal usability principles for educational websites increases productivity and saves time, cost and efforts.

References

- [1]. Prafulla bafna, Shailaja Shirwaikar, Human Computer Interaction-paradigms, process, practices, Advances in computer vision and information technology, part 20, pg 989-998, 2009.
- [2]. Nielsen, Iterative User Interface Design, IEEE Computer Vol. 26, No. 11 (November 1993),
- [3]. Alan Dix, 'Human computer interaction', prentice-Hall, Microsoft Corporation, 2004
- [4]. Lesh, N. MERL, Zaki, M.J., Scalable feature mining for sequential data, Intelligent Systems and their Applications, IEEE (Volume: 15, Issue: 2), 2000
- [5]. Chapter 16: Clustering Algorithms, Edie Rasmussen, University of Pittsburgh
- [6]. Huan Liu and Lei Yu, Toward Integrating Feature Selection Algorithms for Classification and Clustering, IEEE Transactions on Knowledge and Data Engineering, v. 5 n. 6, p. 914-925, 2004
- [7]. Prafulla Bafna¹, Pravin Metkewar^{2*} and Shailaja Shirwaikar^{3#}, "Novel Clustering approach for Feature Selection", American International Journal of Research in Science, Technology, Engineering & Mathematics" (ISSN (Print): 2328-3491, ISSN (Online): 2328-3580, ISSN (CD-ROM): 2328-362) September-November 2014 issue 8 volume
- [8]. Prafulla Bafna et al, Comparative Analysis of Apriori and Improved Apriori Algorithm, International Journal of Emerging Technologies in Computational, and Applied Sciences (IJETCAS), Issue 7 Volume 2, 135-143, 2014