IJournals: International Journal of Software & Hardware Research in Engineering
ISSN-2347-4890
Volume 3 Issue 8 August, 2015

Improve the analysis of User Navigation using Precision, Recall and F-score

Aastha Barche
PG scholar, CSE department,
Department,
SVITS, Indore, MP.
aastha126@gmail.com

Mr. Anand Sing Rajawat
Assistant professor, CSE

SVITS, Indore, MP.
anandsrajawat@gmail.com

ABSTRACT

Generally the website's user navigates the pages and server cannot fulfil the request that is generated by user or we can say that sometimes server cannot respond required pages of user. The basic need here graphically representation of user's interaction to different web browsers by using User's Browser Analysis Chart, PRECISION Chart, RECALL Chart and F-SCORE Chart. *We face many problems related to the web site navigation and some problems are mention below-*

(1)WebPages information is not designed graphically.

(2)We cannot find the proper way of no. of users.

(3)It's difficult to find out user's need on web site.

In this paper we are going to analysis about the web site navigation .Here we predict the users demand and improve the effectiveness of user web site navigation in graphical way with the help of a mathematical model. In this analysis we address how to improve a web site without introducing substantial changes.

We are going to predict the user's demand or we can say that find out exact no. of user. We know that, already various methods have been proposed to improve navigation of a web site using navigation data and to predict a new unpredictable method is difficult.

General Terms
Navigation, Mathematical model programming, User efficiency

Keywords
Web Mining, Precision and recall

INTRODUCTION

With the help of web structure design and different graphical pattern, we improve the efficiency of the web site navigation with controlling whole web structure. It's very difficult for the developer to clearly understand the user's need, so the developer organizes the web pages based on his own judgement. But we know that the web site effectiveness calculate on the basis of user's satisfaction in place of the developer's understanding. Users interest and their relationship in navigation is more important rather than the developer, so the web site administrators find out the problem and then improve that problem and make their web sites more easily.

It's very important for the administrators to know all information related to the users and their need, because if administrators know all things about user and he provide better web pages or find the interesting pages for the user. We all know that www (World Wide Web) is the source of useful information or data [1][2]. WWW (World Wide Web) develops going to improve the efficiency of web site and reduce the size, traffic and difficulty of web site.

We solve our all the problems related to the web site navigation with the help of design web manager by which we can elaborate the exact no of users and also show all users in graphical form. For performing graphical representation number of users and exact navigation with reference to total navigations we *used* User's Browser Analysis Chart, PRECISION Chart, RECALL Chart and F-SCORE Chart.

Web usage mining, the analysis of user navigation paths through web sites, is a common technique for evaluating site designs or adaptive hypermedia techniques. However, often it is hard to relate aggregated clusters or measures to actual user navigation behaviour. Several approaches to reduce the amount of data, such as clustering, binding, filtering and

hierarchization of documents or usage patterns, have been applied in various systems [3]. In this paper we presented the Navigation Visualizer, a web usage analysis tool that combines the mathematical, high level approach of web usage mining with interactive graph based visualizations.

RELATED WORK

The growing body of literature on web search engine evaluation is purely descriptive in nature and has little consistency. Scoville (1996) surveyed a wide range of web search engines for examining the relevance of documents retrievable through them [4]. The first ten hits evaluated for precision have shown Excite, Infoseek and Lycos superior. Leighton (1996) evaluated the precision of Infoseek, Lycos, WebCrawler and WWW Worm using eight reference questions and rated Lycos and Infoseek higher. Ding and Marchionini (1996) investigated Infoseek, Lycos and Open Text for precision, duplication and degree of overlap using five complex queries [5]. The first twenty hits assessed for precision show that the best results are obtained from Lycos and Open Text. Leighton and Srivastava (1997) searched fifteen queries on AltaVista, Excite, HotBot, Infoseek and Lycos taking the first twenty hits for evaluation of precision [6]. Chu and Rosenthal (1996) have investigated AltaVista, Excite and Lycos for their search capabilities and precision. The authors have used ten search queries of varying complexity by evaluating the first ten results for relevance assessment and revealed that AltaVista outperformed Excite and Lycos both in search facilities and retrieval performance [7]. Clarke and Willett(1997) searched thirty queries of varying nature on AltaVista, Excite and Lycos and obtained best results in terms of precision, recall and coverage from AltaVista.Bar-Ilan (1998) investigated six search engines using a single query "Erdos". All 6,681 retrieved documents examined for precision, overlap and an estimated recall report that no search engine has high recall [8][9].

Website navigation has been based on the most important design features across many domains, including finance, e-commerce, entertainment, education, government, and medical, from human computer interaction's perspective, interface, including graphical design, layout design and usability analysis also play an important role in website design. System design including hardware design, cache scheduling, affect the website performance [10]. Besides, structure design, including hyperlinks configuration and information structural design, has a great effect on the website navigation. We know that previous research has taken a different place in history. Navigation is one of very common tasks of web search activity, and information finding task accounts for 25% of search activity. The information make suitable on shortest path will not change by adding a link. Literature shows that adding shortcut may reduce average number of searching steps. However, as number of link increases, the complexity of graph increases and the search time may not decrease by adding shortcuts.

Old mathematical programming model is used to improve the navigation effective for a website while minimizing changes to its current structure; an important issue has not been examined in the literature. Model is particularly appropriate for informational websites whose contents are relatively stable over time. It improves a website rather than reorganizes it and hence is suitable for website maintenance on a progressive basis. The tests on a real website showed that our model could provide significant improvements to user navigation by adding only few new links. Optimal solutions were quickly obtained, suggesting that the model is very effective for real world websites. In addition, we have tested the mathematical programming model with a number of synthetic data sets that are much larger than the largest data set considered in related studies as well as the real data set. The mathematical programming model was observed to scale up very well, optimally solving large-sized problems in a few seconds in most cases on a desktop personal computer.

In old paper the Navigation Visualiser, a web Usage analysis tool that combines the mathematical, high level approach of web usage mining with interactive graph based visualizations. It facilitates tracing and understanding user actions, which would be harder to do with either one of the approaches. Furthermore, the Navigation Visualiser provides means for pre-processing complex user data for further analysis in statistical packages.

This study aims to improve Web navigation efficiency by reorganizing Web structure. To avoid users to lose their orientation, structure stability is taken into consideration. This analysis proposes a mathematical programming method to reorganize Web structure in order to achieve better navigation efficiency. User can specify the requirements and how stable the website structure should be. Navigation efficiency is defined systematically and an e-banking example is given to illustrate how the method works. This study has the advantage of improving navigation efficiency mathematically and relieving the designer of tedious chore to modify the structure in transformation. If some constraint, e.g. structural stability constraint, is relaxed, there is still room for improvement.

Some improvements are significant but will saturate or even decrease as after adding a certain number of links. Research Contribution and Managerial Implications We proposed a quantitative method to improve navigation efficiency systematically. From problem solving, we identify some potential important research areas that has not been explored or paid attention to. For the website designer in the company, it is very important to evaluate the efficiency quantitatively Moreover, to improve a website should be put on schedule after the assessment if the result is not satisfactory.

PROPOSED METHODOLOGY

In this paper, we proposed a mathematical programming model to improve the navigation of a website to understand current structure.

This model is used for improve the web site with the help of graphical patterns or we can say that this mathematical model is very useful for maintenance of a web site on the server side.

This kind of model provides good controlling on website navigation only by adding new features of graphical pattern.

We analysis this model is very effective for the website navigation and understand the traffic.

Precision

Precision or positive predictive value is the fraction of retrieved instances that are relevant under pattern recognition and information retrieval binary classification system. In the field of information retrieval, precision is the fraction of retrieved documents that are relevant to the query:

$$\text{precision} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}$$

Lets we take example of a program for recognizing dogs in scenes from a video identifies 7 dogs in a scene containing 9 dogs and some cats. If 4 of the identifications are correct, but 3 are actually cats, the program's precision is 4/7.

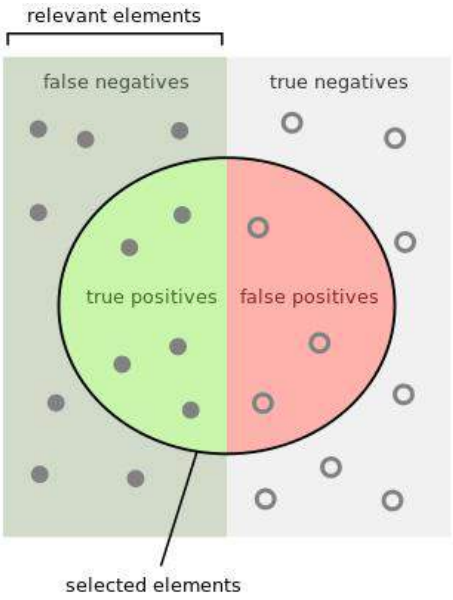

Fig. 1

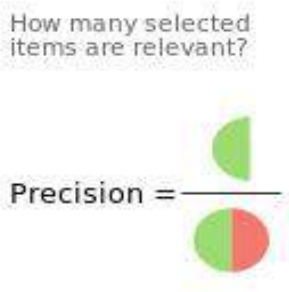

Fig. 2

Recall

Recall or sensitivity is the fraction of relevant instances that are retrieved under pattern recognition and information retrieval binary classification system. Recall in information retrieval is the fraction of the documents that are relevant to the query that are successfully retrieved.

$$\text{recall} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{relevant documents}\}|}$$

Lets we take example of a program for recognizing dogs in scenes from a video identifies 7 dogs in a scene containing 9 dogs and some cats. If 4 of the identifications are correct, but 3 are actually cats, the program's recall is 4/9.

Fig. 3

Relation between Precision and Recall

Precision and Recall are based on an understanding and measure of relevance.

Often, there is an inverse relationship between precision and recall, where it is possible to increase one at the cost of reducing the other.

In a classification task, the precision for a class is the number of true positives (i.e. the number of items correctly labelled as belonging to the positive class) divided by the total number of elements labelled as belonging to the positive class (i.e. the sum of true positives and false positives, which are items incorrectly labelled as belonging to the class). Recall in this context is defined as the number of true positives divided by the total number of elements that actually belong to the positive class (i.e. the sum of true positives and false negatives, which are items which were not labelled as belonging to the positive class but should have been).

Precision can be seen as a measure of exactness or quality, whereas recall is a measure of completeness or quantity.

In simple terms, high precision means that an algorithm returned substantially more relevant results than irrelevant, while high recall means that an algorithm returned most of the relevant results.

F-Score

F-Score is a Harmonic mean of Precision and Recall.

If no entity is left out without assigning label to it then

Precision = Recall = F-Score

F-Score = 2 (P*R) / (P + R)

Where P= Precision and R= Recall

COMPARISON TABLE

COMPARISION TABLE					
OLD ART ALGORITHM			IMPLEMENTED ART ALGORITHM		
PRECISION	RECALL	F-SCORE	PRECISION	RECALL	F-SCORE
0.28	0.27	0.284	0.875	0.8293839	0.8515815

CONCLUSION

In the next step of navigation we researched in this field by using User's Browser Analysis Chart, PRECISION Chart, RECALL Chart and F-SCORE Chart. These charts maintain the server by providing well information of user and webpage hitting ratio. Different Graphs perform well known pattern that server's performance become superior for serving the web pages.

REFERENCES

[1] Bar-Ilan, J. (1998). On the overlap, the precision and estimated recall of search engines: A case study of the query "Erdos". *Scientometrics*, 42 (2), 207-208.

[2] Chu, H., & Rosenthal, M. (1996). Search engines for the World Wide Web: a comparative study and evaluation methodology. In: *Proceedings of the ASIS 1996 Annual Conference*, October, 33, 127-35. Retrieved August 19, 2003 from http://www.asis.org/annual-96/ElectronicProceedings/chu.html

[3] Clarke, S., & Willett, P. (1997). Estimating the recall performance of search engines. *ASLIB Proceedings*, 49 (7), 184-189.

[4] Ding, W., & Marchionini, G. (1996). A comparative study of the Web search service performance. In: *Proceedings of the ASIS 1996 Annual Conference*, October, 33, 136-142.

[5] Leighton, H. (1996, June 25). Performance of four WWW index services, Lycos, Infoseek, Webcrawler and WWW Worm. Retrieved June 10, 2005 from http://www.winona.edu/library/webind.htm

[6] Leighton, H., & Srivastava, J. (1997). Precision among WWW search services (search engines): AltaVista, Excite, HotBot, Infoseek and Lycos. Retrieved June 11, 2005 from http://www.winona.edu/library/webind2.htm

[7] Library of Congress (2003). *Library of Congress Subject Headings* (vol.s 1-5). Washington: Library of Congress, Cataloging Distribution Service.

[8] Modi, G. (1996). Searching the Web for gigabucks. *New Scientist*, 150 (2024), 36-40.

[9] Oppenheiem, C., Moris, A, Mcknight, C., & Lowley, S. (2000). The evaluation of WWW search engines. *Journal of documentation*, 56 (2), 190-211.

[10] Schlichting, C., & Nilsen, E. (1996). Signal detection analysis of WWW search engines. Retrieved September 15, 2003 from http://www.microsoft.com/usability/webconf/schlichting/schlichting.htm

