

Neural Network based Approach for Predicting User Satisfaction with Search Engine

Sunita Yadav
Amity Institute of Information
Technology
Amity University, Noida

Dr. Om Prakash Sangwan
School of Information &
Communication Technology
Gautam Budha University
Greater Noida

ABSTRACT

Success of a search engine is measured by the satisfaction of its users. Finding user expectation can be a better step for improved user satisfaction. In this paper we have proposed a neural network based approach for predicting user satisfaction with search engine. Our work is divided in two parts. Part I investigates user expectations towards search engine for their information need. In Part II we proposed an Artificial Neural Network (ANN) model for predicting User Satisfaction. In our work we have analyzed the major factors affecting user satisfaction with search engine and find out the importance /priority value of these factors based on a survey conducted on 100 search engine users of different profiles with 5-10 years of experience using search engines for their information needs like study material, entertainment, research, day to day problem solution etc. In the present work we have identified four major factors namely up-to-date information, search result relevancy, response time and reliability, contributing to the user satisfaction and developed an ANN model which predicts satisfaction results with a reasonable degree of accuracy.

General Terms

Search Engine, User Satisfaction, Artificial Neural Network

Keywords

Search Engine, up-to-date information, search result relevancy, Response Time, Weight Value, Reliability, Freshness, ANN and user satisfaction engines.

1. INTRODUCTION

User satisfaction is one of the most important factors to measure the success of search engines. It is the key variable in the search process where the information that has been obtained is evaluated and utilized by the information seeker. Therefore finding the factors which are very important for the user while searching information on the web using search engine can be a better step towards improved user satisfaction.

User satisfaction with search engines is based upon several factors. It is proved to have a strong relationship with relevance, freshness, presentation of search results. It is understood that greater satisfaction can be achieved by providing user friendly interface, fast response time, good coverage of Web. Further, diversity of content and discovery of new pages/content will play an important role towards user satisfaction. User expectation, behavior and psychology can also affect the users' satisfaction. There are several human factors which may contribute to the overall level of user satisfaction.

With the drastic growth of the World Wide Web, artificial neural network learning algorithms have been used in Web-related applications such as Web usage analysis, Web searching, and Web page clustering [7, 15, 17]. Most traditional neural network systems only rely on the available input-output examples [24].

Artificial neural network learning algorithms have been functional in various applications such as classification, clustering, and pattern recognition by modeling the human neural system. These applications have been studied and tested in various areas including engineering, medicine and finance among others. Neural network models have also been widely used in the area of information retrieval and text mining, such as text classification, text clustering and collaborative filtering.

The paper is organized as follows. Major factors affecting user satisfaction with search engine are discussed in section 2. Survey on user satisfaction is evaluated in section 3. Proposed methodology for estimating user satisfaction with search engine is presented in section 4. Experimental results are discussed in section 5. We draw the conclusion with future directions in section 6.

2. MAJOR FACTORS AFFECTING USER SATISFACTION

User satisfaction can be considered as one of the indicators of user's achievement of information need. Besides this user friendly interface, fast response time, good coverage of Web, diversity of content, discovery of new pages/content, interactivity and other factors have significant impact on user satisfaction. Short response time, easy navigation, appropriate use of graphics and high degree of interactivity are more likely to attract users [20]. There are number of factors that may affect user satisfaction as user relies on search engine for information need:

- a. **User friendly Interface:** Easy to use interface is likely to attract user. User interface should be designed to stimulate system use and increase productivity while reducing learning time.
- b. **Extra Services:** Spelling suggestions, current news, images, local results, stock chart etc. are examples of elements, that modern web search engine return to the user.
- c. **Long Tail Keyword Sets:** Search Engines ability to provide useful information for the rare queries too.

- d. **Response Time:** Time taken by a search engine to provide search engine results when user submits a query for information.
- e. **Presentation:** Order and number of search engine results on a page.
- f. **Coverage:** Amount of content on the web which is searchable by the search engines.
- g. **Diversity:** Ability of the search engine to handle ambiguous queries.
- h. **Discovery & Latency:** Ability and time taken by the search engine to provide latest content published on the web.
- i. **Relevance:** How much the retrieved results meet the user's information need.
- j. **Freshness:** Ability of the search engine to handle ever-changing content of the web.
- k. **Geographic Information:** Ability of search engine to identify implicit geographic information and produce results accordingly.
- l. **Reliability:** Links provided by the search engine in the results are reliable or not.
- m. **Availability:** Whether the search engine site is available every time the user wants to search for information.
- n. **Help & Support:** If a user wants some help regarding the use of search engine tool for information extraction, is it provided by the search engine or not?
- o. **Future Evolution:** Does the search engine provide updates about its future evolutions?

3. SURVEY ON USER SATISFACTION

3.1 Survey Technique

A survey was conducted from a sample size of 100 diverse search engine users to identify the importance of the above mentioned factors for users as they query search engines. In a written questionnaire, people were asked to rank the factors according to the importance they them as they select a search engine for searching information.

Survey sample consisted of heterogeneous group of participants consisting of research scholars, corporate employees, faculties, tech-savvy people and some naive users.

3.2 Survey Result/Behavior based Evaluation

After carefully studying and scrutinizing the data collected from the questionnaire, the factors affecting user satisfaction are categorized as primary(P) and secondary(S).

Table 1: Weight Value for Search Engine Selection Factors

P-Primary, S-Secondary	Factors	Normalized Score
P1	Response Time	0.037
P2	Up-to-date Information	0.027

P3	Reliability	0.037
P4	Search Result relevancy	0.034
S1	Presentation of Search Results	0.088
S2	'Look and feel', friendliness & easiness of User Interface	0.076
S3	Ability to handle different interpretations of a query(Diversity)	0.073
S4	Ability to provide results based on geographical requirements	0.086
S5	Extra services like news, stock chart etc.	0.113
S6	Ability of providing useful information for the rare queries too	0.073
S7	Ability to provide all the information which is available on the web	0.076
S8	Availability	0.067
S9	Help & Support' features for users	0.104
S10	Search Engine's future evolution updates	0.110

Table 1 show that up-to-date information is the most preferred criteria for search engine selection while users gave least preference to future evolution updates. Thus up-to-date information, Search Result relevancy, Response Time and Reliability are the most important factors for satisfying the maximum users.

4. NEURAL NETWORK METHODOLOGY

Artificial Neural Networks (ANNs) have the ability to model complex non-linear relationships and are capable of approximating any measurable function. ANNs have several features which make them attractive prospects for solving pattern recognition tasks without having to build an explicit model of the system. In ANN modeling, we determine a pattern of connections (architecture), a method of determining interconnection weights w (training algorithm) and transformation function. The number of input unit and the number of output unit are problem dependent. In our work we have one output unit i.e. user satisfaction and four input units namely Response Time, Up-to-date information, Search Result Relevancy and Reliability.

In the present work, we have identified factors contributing to user satisfaction and proposed Artificial Neural Network (ANN) model for evaluating it. Neural networks have been established to be an effective tool for pattern classification and clustering [5, 12]. The major advantage of ANN is that it learns by example and has the ability to generalize from its training data to other data. ANN's are relatively inexpensive to build and train [6].

Generally, the neural network models have been used for engineering predictions, economic predictions, data mining and medical diagnoses [22]. The major studies reporting the

application of neural networks in web searching have been carried out in web content mining, personalization, clustering and web content classification and result relevancy.

Web mining is the use of data mining techniques to retrieve, extract and evaluate information from web documents and services. Web mining typically addresses semi-structured or unstructured data, e.g., web and log files, which are often represented by inaccurate or incomplete information. Thus neural networks are a useful instrument to mine knowledge from such data.

Personalization is main factor for the interactions to a website to cater the user's interests and need. Personalization has become mandatory now a day and requires flexible customization. Web recommendation systems using neural network is one of the successful examples of personalization [16, 9].

The World Wide Web is a very vast, useful and valuable resource, whose full potential cannot be realized until its content is effectively organized and described in a simple way. However, due to the immense size and dynamic content of the web, manual customization or categorization is not a practical solution to this problem. Thus there is a clear demand for automated classification of web content. However it is not possible to increase the accuracy of web classification due to the diversity in the web pages' content. Neural network can be used to filter the data on the web for classification [2, 8, 11, 14, 19, 21, 23].

Normally the search engines search for data in their databases by using the exact words in the user query. As many users cannot articulate the search with the deliberate keywords, search engines fail to return the expected results. It is found that the user usually only go through the first page of the search results; hence, it becomes very important to output the best possible result to the user through their individual perspective. Search engines which are aided with neural network can provide more relevant result to the users' query [3, 4, and 18].

A news web page classification method (WPCM) using a neural network with inputs obtained by both the principal components and class profile based features was proposed by Selamat et al. [1]. Ko-Kang Chu et al. [10] applied hybrid training neural network to Web-based applications.

Based on the ANN research survey, we adopted ANN to predict the satisfaction of user with the search engine.

5. EXPERIMENTAL DESIGN

The normalized score of the data collected from user survey acts as input data sets to our proposed Neural Network model. It is known that NN model can be used to model any arbitrary input-output mapping and are capable of approximating any measuring function. So NN should be able to model the measurement of user satisfaction also.

From the survey, we identified the following four major factors on which the satisfaction of the user with search engine depends:

- i. Response Time,
- ii. Up-to-date information,
- iii. Reliability
- iv. Search Result Relevancy

i. Response Time

Response Time is defined as the time taken by search engine to provide results when user submits a query for information. For the present work, we categorize it from slow to fast.

ii. Up-to-date information

Up-to-date information is defined as the ability to provide latest content published on the web and to handle ever-changing content of the web. For the proposed experimentation, we assume that Up-to-date factor ranging from low, medium and high.

iii. Reliability

Reliability is defined as the links provided by the search engine in the results are reliable or not. For the proposed experimentation, we assume reliability ranging from low, medium and high.

iv. Search Result Relevancy

Search Result Relevancy is defined as how much the retrieved results meet user's information need. For present work, we categorize relevancy into low, medium and high categories. For the proposed experimentation, we assume that relevancy ranging from low, medium, and high.

5.1 ANN Modeling

In our study, back propagation algorithm is used. The model is trained with four inputs namely: Response Time, Up-to-date information, Search Result Relevancy and Reliability with one output i.e. user satisfaction with search engine.

The network used in this work belongs to multilayer feed forward network. The summary of ANN used in this study is shown in Table 2.

Table2: Neural Network Model Summary

Architecture	
No. of Neurons at Layer 1 (Hidden Layer)	25
No. of Neurons at Layer 2 (Output Layer)	01
Input Units	04
Output Units	01
Training	
Training Algorithm	Feed Forward Back Propagation
Training Function	Trainbr
Transfer Function	Tansig

5.2 Training

Network is trained by adjusting different weights of neurons and hidden layers, so that the difference between desired output from the network and actual output is minimized. The network learns by finding a vector of connection weights that minimizes the desired error on the training data set. The training data set was consisted of 160 exemplars divided into 12, 43, 64, 29 and 12 cases of very low, low, medium, high and very high categories of user satisfaction respectively. The network is investigated empirically containing different numbers of neurons, i.e., 12, 16, 20, 25, 28 and 32 in hidden layer(s). The network is best trained with four inputs, one output and 25 neurons at the hidden layer as shown in figure 1.

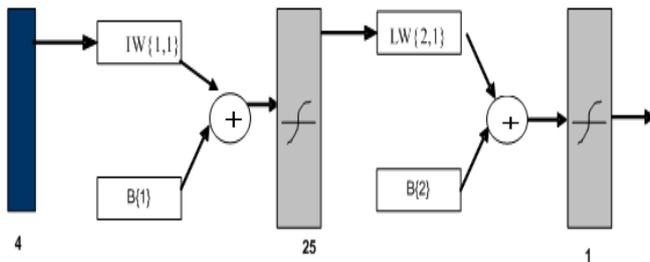


Figure 1: the architecture of the neural network used, 04 inputs, 25 nodes in the hidden layer and 01 in the output layer.

Trainbr function of Feed Forward Back Propagation algorithm is used as training function to train the network as shown in figure 2. Linear transfer function *Tansig* has used as a transfer function for the experimentation. The adaptation learning function selected for this experiment was '*learnqdm*'. We used the transfer functions, '*tan-sigmoid*' in both the layers, which converts the input into net output. Each of the networks was trained with 100 epochs. Other parameters are set at their default values.

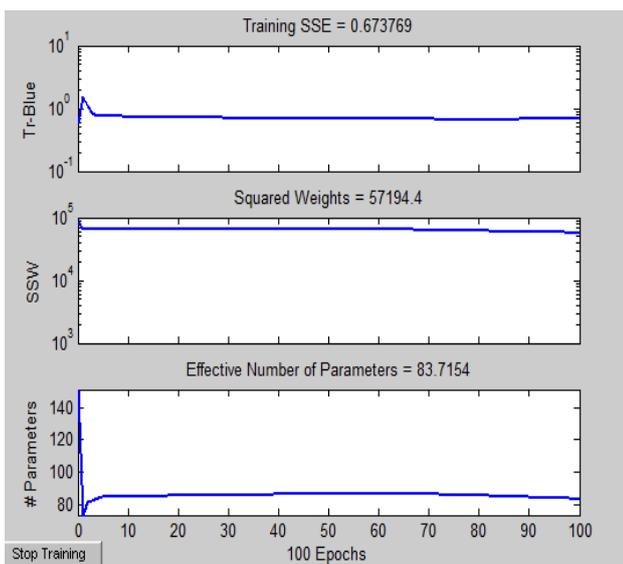


Figure 2: Training with *trainbr* function

5.3 Performance Evaluation

In order to establish the success and adequacy of supervised training for the neural network models, it is necessary to have some quantitative measure of learning. Root Mean Squared Error (RMSE) is an adequate and commonly used error measure. RMSE is a useful measure of how close a network is getting its predictions to its target output values. For successful training, RMSE will decrease significantly in the initial stages of training and converge after a sufficient number of iterations have been completed. It is calculated for use in this study as follows:

$$RMSE = \sqrt{\frac{\sum (Actual\ Value - Estimated\ Value)^2}{n}} \quad (1)$$

Where n is the number of observations.

The optimum performance of the Feed Forward Back Propagation model is found to be by *trainbr* function at 25 neurons with RMSE 0.1481. The accuracy of the obtained results with different neurons is shown in Table 3.

Table 3: Accuracy for the Network

User Satisfaction	
Neurons in the Hidden Layer	RMSE for Trainbr function
12	0.2046
20	0.2154
25	0.1481
28	0.2197
32	0.2087

6. EXPERIMENTAL RESULTS

The goal of our work is to develop a tool for measuring user satisfaction with search engine. Experiments were conducted with an ensemble of networks, starting from network with 12 hidden neurons and going up to 32 neurons to find out which of these networks gives the best performance, keeping the output layer neurons constant in each of the networks i.e. 1 representing one of the five target values. After the training, simulation was done on the network. The network was tested against two test sets (validation set 1 & validation set 2), consisting of 80 exemplars each. Table 4 shows the RMSE results of ANN model evaluated on validation data sets.

Table 4: Validation results of ANN model

User Satisfaction		
Neurons in the Hidden Layer	RMSE (Validation Set 1)	RMSE (Validation Set 2)
25	0.1481	0.1382

Comparison of the experimental results and measured user satisfaction of the two validation sets is shown in figure 3 and 4.

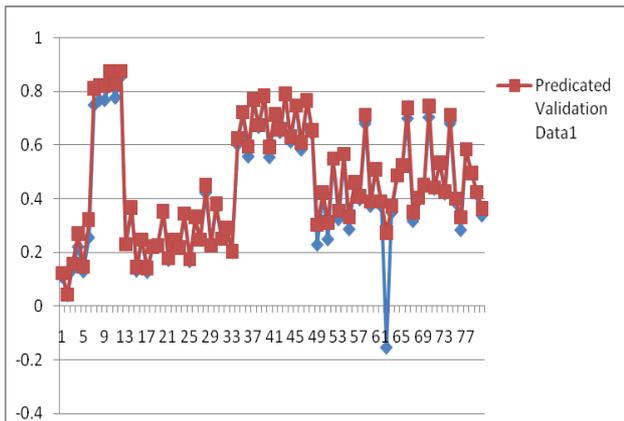


Figure 3: Validation Results (Data Set 1) with 25 neurons

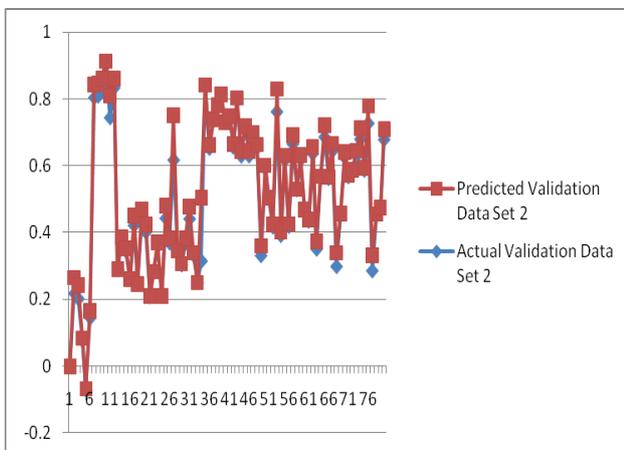


Figure 4: Validation Results (Data Set 2) with 25 neurons

Therefore, from the experimental results, the proposed ANN based approach is suitable for predicting the satisfaction of the user with search engine.

7. CONCLUSION AND FUTURE DIRECTIONS

It is not a single factor upon which user satisfaction with search engine is based but several factors contributes to it. Our study has shown that user satisfaction is related to and affected by several factors. Few of them are discussed in the paper.

Present work adopts artificial neural network based approach to measure user satisfaction with search engine. Network is trained on training data set of 160 inputs by considering different number of hidden neurons for training function namely, *trainbr*, to get the best results. This network is further validated on test set of 80 exemplars by applying the proposed approach on test data. Results obtained with RMSE 0.1481 shows that network is able to predict the user satisfaction with search engine with an acceptable accuracy.

It is possible to explore alternative neural network model like cascade feed forward back propagation model that dynamically build the Neural Network architecture model. However the results may vary depending upon the Neural Network architecture, training algorithm, training function, transfer function and data set.

8. REFERENCES

- [1] Ali Selamat and Sigeru Omatu, "Neural Networks for Web Page Classification Based on Augmented PCA", Proceedings of the International Joint Conference on Neural Networks, 2003, Volume 3, pages 1792 – 1797
- [2] Anagnostopoulos, I. Et al., "Classifying Web pages employing a Probabilistic Neural Network", Proceedings: Software, v151, n3, June 2004, page 139-150.
- [3] Franco Scarselli, Sweah Liang Yong, Markus Hagenbuchner, Ah Chung Tsoi, "Adaptive page ranking with neural networks", Special interest tracks and posters of the 14th international conference on World Wide Web, May 2005, pages 936-937.
- [4] Giovanni Pilato, Salvatore Vitabile, Giorgia Vassallo, Vincenzo Conti, Filippo Sorbello, "A Neural Multi-Agent Based System for Smart Html Pages Retrieval," International Conference on Intelligent Agent Technology (IAT'03), October 2003, pages 233.
- [5] Haykin, S. (2003), Neural Networks, A Comprehensive Foundation, Prentice Hall India.
- [6] Hudson, D. L., Cohen, M. E., 2003. Neural Networks & Artificial Intelligence for Biomedical Engineering, Prentice Hall of India.
- [7] H. Chen, M. Chau, and D. Zeng, "CI Spider: A tool for competitive intelligence on the Web," Decis. Support Syst., vol. 34, no. 1, pp. 1–17, 2002
- [8] Jiu-zhen Liang, "Chinese Web page classification based on self organizing mapping neural networks," Computational Intelligence and Multimedia Applications, ICCIMA, Sep 2003, pages 96-101.
- [9] Kartik Menon and Cihan H. Dagli, "Web Personalization using Neuro-Fuzzy Clustering Algorithms," NAFIPS 2003 22nd International Conference of the North American, IEEE, July 2003, pages 525-529.
- [10] Ko-Kang Chu, Maiga Chang and Yen-Teh Hsia, "A Hybrid Training Mechanism for Applying Neural Networks to Web-based Applications", International Conference on Systems, Man and Cybernetics, 2004, pages 3543-3547.
- [11] Lee, P.Y.; Hui, S.C.; Fong, "Neural Networks for Web Content Filtering", Intelligent Systems, IEEE Volume 17, Issue 5, Sep/Oct 2002, pages 48 – 57.
- [12] Mayrhauser A., Anderson C. and Mraz R. (1995), Using a Neural Network to Predict Test Case Effectiveness, Proceedings of IEEE Aerospace Applications Conference, Snowmass, CO, pp:77-91.
- [13] Mukherjee, A. and Deshpande, J. M., 1995. Neural Network Based Expert Systems for Structural Design, Computers and Structures, Vol. 54, Iss. 3, pp 367-375.
- [14] Ma. Heng "Fast Blocking of Undesirable Web pages on Client PC by discriminating URL using Neural Networks," Expert Systems With Applications, 2007.

- [15] O. Gogan and S. C. Buraga, "The use of neural networks for structural search on Web," presented at the Int. Symp. Syst. Theory—SINTES10, Craiova, Romania, May 25–26, 2000.
- [16] Olfa Nasraoui, Mrudula Pavuluri, "Accurate web recommendations based on profile-specific url- predictor neural networks", Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters WWW Alt. '04, May 2004, pages 300-301.
- [17] P. Batista and M. J. Silva, "Web access mining from an on-line newspaper logs," presented at the 12th Int. Meet. Euro Work. Group Decis. Support Syst. (EWG-DSS 2001), Cascais, Portugal, May 2001.
- [18] Seda Özmutlu, Fatih Çavdur. "Neural network applications for automatic new topic identification." Online Information Review, 2005, pages 34-53.
- [19] Selamat A., & Omatu S. (2003). "Web Page Feature Selection and Classification Using Neural Networks", Information Sciences.
- [20] Sunita Yadav, Omprakash Sangwan, "Standardization Specification of Factors Affecting User Satisfaction with Search Engine", Proceeding of National Conference on Advanced Computing and Communication Technology- ACCT2010, June 2010 pages 909-914.
- [21] Vljajic, N.; Card, H.C., "An Adaptive Neural Network Approach to Hypertext Clustering", Neural Networks, 1999. IJCNN'99. International Joint Conference, Volume 6, Jul 1999 pages 3722-3726.
- [22] Widrow, B., Rumelhart, D. E., and Lehr, M. A., 1994. Neural networks: Applications in industry, business, and science. Communications of the ACM, Vol. 37, pp: 93-105.
- [23] Xing Zhu, Shen Huang, Yong Yu, "Web technologies and applications: Recognizing the relations between Web pages using artificial neural network", Proceedings of the 2003 ACM symposium on Applied computing SAC '03, ACM Press, 2003, pages 1217-1221.
- [24] Y. Jin and B. Sendhoff, "Knowledge incorporation into neural networks from fuzzy rules," Neural Process. Lett., vol. 10, pp. 231–242, 1999.
- [25] Ying Xie, Dheerendranath Mundluru, Vijay V. Raghavan, "Incorporating Agent Based Neural Network Model for Adaptive Meta-Search", Proceedings of the 43rd annual Southeast regional conference, Volume 1, Mar 2005, pages 53-58.