nternational Conference on Advanced Computing Technologies and Applications(ICACTA-2015)

E-commerce website ranking using semantic web mining and neural computing

Neha Verma[a], Dheeraj Malhotra[b], Monica Malhotra[c], Jatinder Singh[d]

[a]Department of Computer Science and Engineering,Punjab Technical University, Kapurthala-144601,Punjab,India
[b]Department of Computer Science and Informatics,University of Kota,Kota 324005,Rajasthan,India
[c]RDIAS,GGS IP University, Delhi-110034, India
[d] KCCollege of Engineering & IT, Nawanshahar-144514, Punjab, India

Abstract

With the acceleration of Internet era, E-commerce industry has grown rapidly and a good E-Commerce website has become indispensable for every commerce based organization. However E-Commerce industry in developing countries like India is still lagging behind to satisfy the challenging and dynamic needs of consumers. Enterprises are deploying new strategies to retain or rebuild relations with their old customers and persistently focusing on new customers. Keeping in mind the thought of blending intelligent web mining with E-Commerce, this research work discusses the design of a Semantic and Neural based E Commerce page ranking algorithm, named SNEC page ranking algorithm and its implementation in the form of website ranking tool that may be well utilized for ranking of E- Commerce websites to assist the customers by finding relevant web sites on the top during their search for buying a specific product as well as businesses to compare their strengths and weaknesses with competitors and hence to improve their profits by providing relevant product at competitive price with consumer friendly services to the customers by better structuring their E Commerce websites.

* Dheeraj Malhotra. Tel.: +91-0-9560375531;
*E-mail address:*dheerajmalhotra@ymail.com

1. Introduction

Ecommerce is fastest growing sector in some of the developed countries over the last 10 years. In order to well satisfy the demands of today's customer, now companies are paying the hard work to maintain their competitive reputation and to bring good revenues, growth, and goodwill to the enterprise. Various research based E-Commerce statistics show that India is also growing to more advanced stage with in E Commerce sector. There are many reasons for sudden growth of this sector in developing countries like India such as high computer literacy rate, busy lifestyle, high income groups, trial and easy exchange policies, feedback availability of products, cash back, credit on purchase and cash on delivery like attractive and reliable policies. One of the growing and associate research area, Intelligent Neural based Web Mining with the features of adaptability and learning from errors is also transiting from laboratory stage to practical application stage and is quite helpful to extract useful information from customer's surfing pattern on the web. With the help of these patterns, we can improve the website structure; enhance availability of information in website, ease of information accessibility to users. In this paper we will discuss about how web mining can be deployed intelligently to get benefit in various dimensions of E-Commerce, which is not only useful for customer but also for data analysts to take various important decisions to better satisfy their organizational needs. Accordingly this paper is divided into various sections. Section II describes related work, Section III explain about the research problem, Section IV illustrates the objectives of research. Section V gives clear understanding of research methodology, proposed and innovative SNEC page ranking algorithm and website priority determination tool followed by graphical analysis of results.

2. Related Work

E-Commerce when supported by semantic and neural based intelligent web mining technology leads to useful patterns for better ranking of E Commerce websites. The objective of such a unique blend is to assist the customer to select the E Commerce website while carrying out online transactions as well as E Commerce organizations while taking critical decisions for optimizing the structure of their websites. However all the old mining techniques of E Commerce website ranking are not efficient enough to satisfy the dynamic needs of today's customer and hence the research in this area is continuously emerging. Zhov Fan Yang, Yong Shi, Bo Wang, Hong Yan [1] discussed Ecommerce website quality and profitability evaluation using two stage DEA model. They focussed on improving E Commerce website quality and profitability based on sub stage efficiency scores. They compared efficiency scores of CCR, BCC and KH models. Various criteria used by these models to judge the quality of E commerce website may be utilized in ranking them. OguzMustapasa, DilekKarahoca[2] discussed about applications of the combination of data mining with semantic web since they both are addressing each other to fulfil the aim of automated examining of large amount of data from websites and helps in discovering and obtaining meaningful results which may be used for E Commerce website restructuring and ranking.WeigangZuo, QingyiHua[3] suggested various applications of data mining with E-Commerce. They further described about various data mining techniques which has contributed to determine customer behavior and feedback, which in turn is helpful for website structure optimization. Yanduo Zhao [4] discussed about E-Commerce websites, they produce large amount of data and important information which is hidden. This hidden information upon retrieval may be used for effective website restructuring as well as to place its web pages in correct sequence which will improve the rank of website. Shenzihao,Wanghui [5] suggested architecture of a model using web mining with E-Commerce and its application area on website structure optimization, personalized recommendation, business intelligence and network security etc. were discussed. They also discussed about the relevance of constant watch on user visiting patterns which can be helpful for website structure optimization..Dheeraj Malhotra, Neha Verma [8] well described web dictionary based page rank determination algorithm. The proposed algorithm determine relevancy of web page using page content and time spent by previous user. The objective is to improve the time and space complexities of search engine algorithms while searching in their huge web databases without compromising with user experience. This algorithm is further extended into improved and innovative SNEC page rank algorithm by incorporation of semantic web and neural computing in our present research work. Shuo Wang, KaiyingXu, Yong Zhang, Fei Li [9] used an approach for personalization and optimization of search engine

which is based on back propagation neural networks. This method is based on feedback from user which may be explicit or implicit. They also explained various methods for obtaining training samples of information retrieval. Back propagation concept of neural networks may be used to implement the unbiased ranking model for E Commerce websites. LI Yaolin, ZHONG Yanhua, NIE Shuzhi [10] suggested quantum self-organizing neural network can better cluster users for dynamic generation of personalized web pages for different categories of users. The proposed model has strong generalization capability and may be used for correct ranking of web sites by profiling user navigation patterns as is discussed in the design of proposed system in our present research work.

3. Research Problem

The exponential growth in the web and E-Commerce industry has been unhidden. In the absence of a concept such as web catalogue, customer is dependent on search engines to find suitable E Commerce web site to purchase a product but some of the common problems associated with search engines include syntactic search of resources which means search engines implement matching process in terms of frequency count, proximity etc between user search query and candidate web page. This syntactic matching lack semantics as a result the product queries which can be interpreted in various contexts are likely to produce wrong results and leads to abundance or scarcity problem and user usually end up with thousands or even more links and sometimes not even a single link in the output of search engine. Moreover the page ranking provided by most of the popular search engines is highly unreliable and is hugely impacted by money making businesses such as SEO which tend to show the pages of their interest on top irrespective of their content, credibility and degree of relevancy to the requirement of customer. As a result customer is not able to find the relevant and genuine product at best price easily. For instance, some of the E-Commerce websites listed usually on the top of search engines results corresponding to E Commerce product query are selling products without taking authorization from the manufacturer of the product at unreasonable prices leading to customer sufferings while taking warranty and guarantee services from manufacturer of the product. One of the reason for such problems include generic design of search engines to understand the customer intension and another reason is that in the absence of back-propagation of errors or feedback mechanism, retrieval algorithms leads to biased ranking which usually end up in making only top ranked pages more popular as user also typically look at only the first few pages of search results while interacting with search engine. Hence there is an urgent requirement to focus on assisting the customer to take informed and intelligent decision to select an appropriate E Commerce website while carrying out online transactions.

4. Objectives of Research

The overall objective of this research work is to improve the E Commerce web site ranking process by developing a *SNEC Page Ranking* algorithm and its implementation in the form of an automated tool to assist the customer while carrying out E-Commerce transactions as well as to assist the E-Commerce website owner to optimize the structure of his website so as to take lead over its competitor websites. In this research work, a semantic and neural based mathematical approach to deal with various ranking problems is discussed and this intelligent ranking algorithm will optimize the use of web dictionary, previously spend time statistic, recommendation engine for incorporation of semantic capabilities and back propagation based neural network to learn from errors and hence to implement the unbiased ranking process. This algorithm may be implemented as an *Intelligent Meta Search Engine*.

5. Research Methodology

The extracted knowledge from semantic and neural based web mining techniques can be used to better rank the E Commerce web sites, attracting new customers, enhance resilience time on websites for existing customers etc. In this research work, we first process the retrieved candidate web page from search engine or manually entered web URL using *Profiling and Dictionary Implementation Module* to implement web log preprocessing in order to remove incomplete entries, data cleaning, removal of stem words, User navigation profiling and finally implementation of a

web dictionary which consist of only those words from candidate web page which are relevant as per length of each of the constituent word of user search query for a specific E Commerce product. This web dictionary and candidate web page is further passed to *Content Priority Module* to apply Web Content mining to check for relevancy of web page and hence to determine its priority as well as the removal of irrelevant web page with respect to specified E Commerce product to be searched. After that web page would be passed to *Time Spent Priority Module* which will determine priority of candidate web page using time stamp of the creation of candidate web page as well as previous user time spent statistic who searched for the same product. This module will assign high priority to those web pages on which previous user spend more time while carrying out E Commerce transaction and are also recently created. Now the web page is passed to *Semantics Recommendation Module* which in turn is responsible for identification of user Session from various navigation pattern profiles using Longest Common Subsequence algorithm and determination of ontology class to avoid wrong interpretation of user search query. The overall priority of E Commerce web site is finally determined by using supervised back-propagation neural network. This network processes candidate E Commerce web site at various layers i.e. Input layer, Hidden layer and output layer. Input layer accepts five inputs i.e. content priority, Time spent priority, E Commerce user explicit and implicit feedback about the candidate web site, recommendation semantics and biased input. Initially it will assign random weights to all of these input links. The actual output of the network is compared to threshold value set by human volunteer. This may result into an erroneous priority followed by determination of error margin between expected and produced priority output. The weights of input links are gradually adjusted via feedback of error margins from output to input layer, until the correct output is produced. Fine tuning the weights in this way has the effect of implementation of supervised learning by the network and hence to assist in determination of correct priority of each of the candidate E Commerce web site with respect to user searched product. The simplified neural design of proposed system is shown in Fig.1 and the overall design of the proposed system is shown in Fig.2.
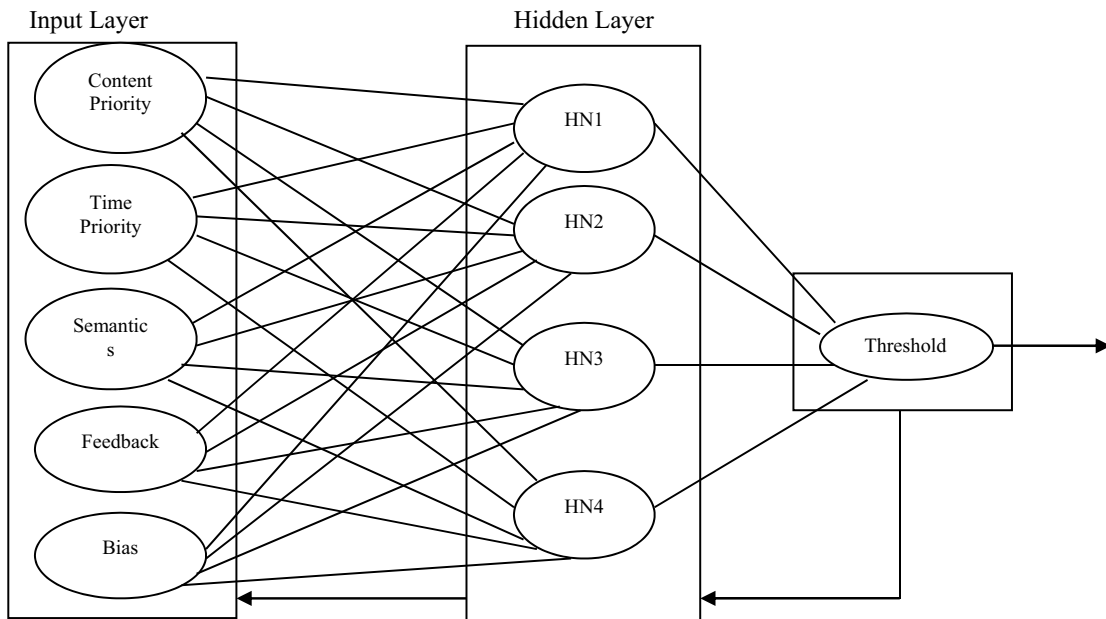

Fig.1. Back propagation supervised neural design of proposed system
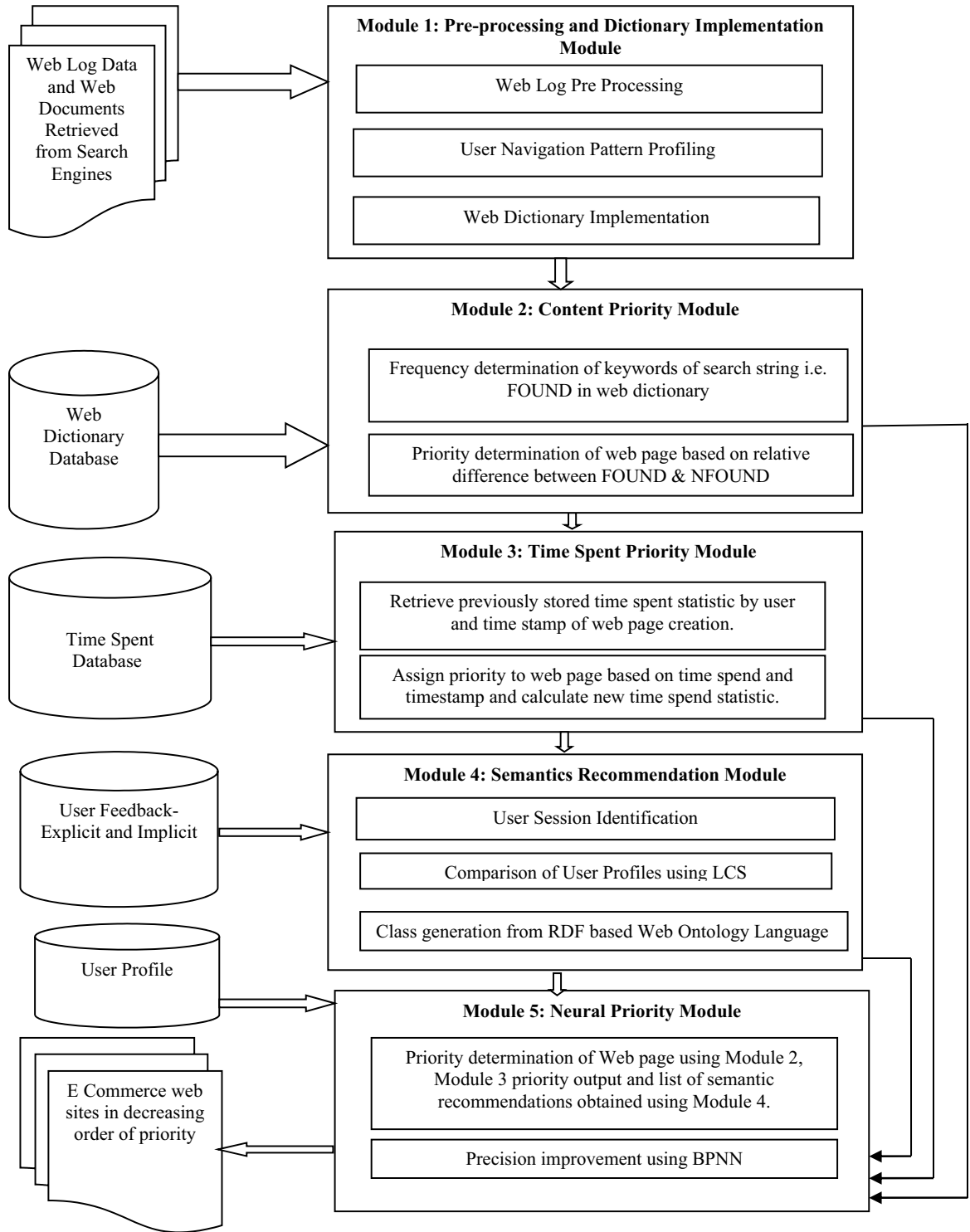
Fig.2. Design of proposed system

5.1 SNEC Page Ranking Algorithm

Nomenclature

SNEC: Semantic and Neural based E Commerce Page Ranking Algorithm

S_i: User search string for E Commerce product.
Min: Minimum length of any of the keyword in S_i.
Max: Maximum length of any of the keyword in S_i.
W_I: Specific keyword in search phrase.
Dp: E Commerce web document to be scanned.
WD_p: Web dictionary corresponding to P_{th} Web document.
DW: Document word.
T_p: Average time spent by previous visitors.
T_s: Time stamp of creation of web page.
Found: Frequency of number of keywords in S_i found in E Commerce web site to be ranked.
Nfound: Frequency of number of keywords in Si not found in E Commerce web site to be ranked.
tan Ø: Linear activation function for training of neural network.
WT_i: Weight of Input Synapses

- **Module 1**

Step 1: Accept search string from user.

Step2: Remove Stem words from search string.

Step3: Record navigation sequence pattern in user profile database.

Step 4: Search the Web documents (say m in number) using search engine.

Step 5: Split the string into various words $W_1, W_2, \ldots\ldots, W_n$.

Step 6: Determine the minimum and maximum length among the various words of search phrase

    min := Strlen(W_1), max := Strlen(W_1)

    for i = 2 to n do

    if min>Strlen(W_i) then

    min: = Strlen(W_i)

    if max<Strlen (W_i) then

    max: = Strlen(W_i)

Step7: Initialize Ti for each document as 0.

Step 8: Search the time database of tool using keywords entered by user and search for the same documents as given by search engine in previous step to retrieve T_i.

Step 9: Preprocess each Web document D_j in dictionary form WD_j allowing only those words DW_k from D_j which satisfies the condition min >= Strlen (DW_k) <= max.

- **Module 2**

Step10: for p=1 to m do

Initialize found$_p$: =0 and nfound$_p$: =0

If W_p found in WD_p then

$found_p := found_p + 1$

Elsen$found_p := nfound_p + 1$

Step 11: Eliminate all those web pages where $nfound_p > found_p$.

- **Module 3**

Step 12: Determine timestamp T_S of creation of web page.

Step 13: On start of user session, determine t_p which is session duration of current page and determine new value of T_p as follows:

If $T_p = 0$ then $T_p = t_p$

Else $T_p = (T_p + t_p)/2$

Step 14: Assign priority high to web page if T_S is low and T_p is high.

Step 15: Update the time database of tool with keywords, page address and T_p.

- **Module 4**

Step 16: Identify navigation session by comparing user search query with each of the search query present in user profile database as

LCS $[i,j] = 0$, if $i=0$ or $j=0$ **OR**

LCS$[i,j] = $ LCS$[i-1,j-1]+1$, if $i,j <> 0$ and $S1_i = S2_j$ **OR**

LCS$[i,j] = $ max(LCS$[i-1,j]$, LCS$[i,j-1]$), if $i,j > 0$ and $S1i <> S2j$

Step 17: Generate class using Web Ontology Language.

- **Module 5**

Step 18: Normalize all the priority inputs from module 2, 3 and 4

Step 19: Train the network using various set of inputs and outputs with linear activation function as

$$\{O\} = \tan \emptyset \ \{I\}$$

Step 20: Use Sigmoidal function for output evaluation in Hidden and Output Layers as: $\{O\} = [(1 / (1+e^{-I})]$ and

Summation function as $\sum(I_1WT_1 + I_2WT_2 + I_3WT_3 + I_4WT_4 + I_5WT_5 + B)$

Step 21: Determine error rate to adjust weights of synapses using supervised learning of BPNN algorithm.

Step 22: Display all the retrieved web pages in decreasing order of their correct priority ranking.

5.2 Website Priority Determination Tool

SNEC algorithm discussed in this research work leads to implementation of Website Priority Determination Tool using ASP.NET framework. The interface of tool allows comparison of at most six E Commerce web sites using dropdown box and search box to specify search string of specific product. Tool will allow entering as many URL's of websites as selected in drop down box .After clicking the comparison button, tool will assign priority to each of the candidate web site based on the calculation of content priority module, time spent priority module, recommendation module and neural priority module. Web site priority determination tool is shown in Fig.3.
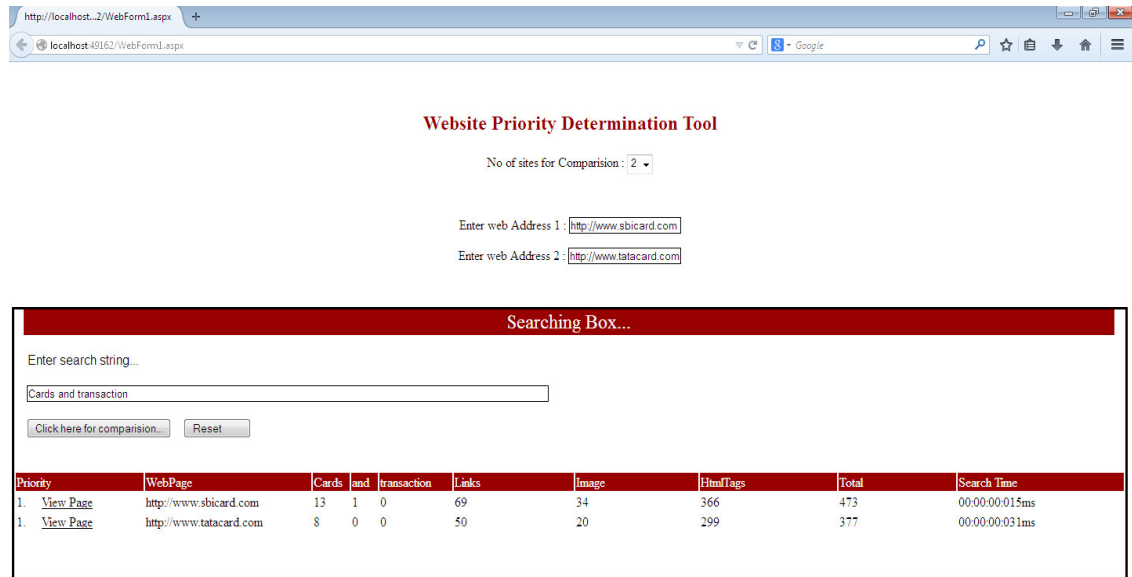
Fig.3. Website priority determination tool

5.3 Graphical Analysis of Results

The effectiveness of website priority tool is dependent on its capability to determine the correct position of a E Commerce webpage for a given product search query. The evaluation method of this research work is based on measurement of precision and coverage. Coverage refers to capability of recommendation system to select all the E commerce websites likely to be visited by the user to search for desired product. All the transactions in the test set are divided into two parts. The first part is used for recommendation based on semantics of the query and the other part is used for evaluation of the recommendation. This research work splits the session logs into the first 80% to be compared by tool and remaining 20% to be produced by the recommendation module to check for coverage. Precision refer to the measurement of number of relevant recommendation with respect to total number of recommendations. Precision is measured at Y metric, which is being denoted by $P(Y)$ and is plotted on Y axis. For a given query, $P(Y)$ reports how many fraction of results that are labelled as relevant are reported in the top Y results. Here the rank provided by Tool and Google is compared with Human mentioned threshold to verify the relevancy and finally the difference in comparison of precision of Tool and Google is plotted in Fig.4 for the same E Commerce product search query. Here it is observed that initially precision of search engine is high. However the precision of tool improves with repeated usage of Tool because of the fact that the neural network implemented in SNEC algorithm will back

propagate the errors to correctly adjust the weights on input links. This simply shows that semantic based back propagation neural network possess the learning capabilities from errors and helps in correct ranking of E Commerce websites.
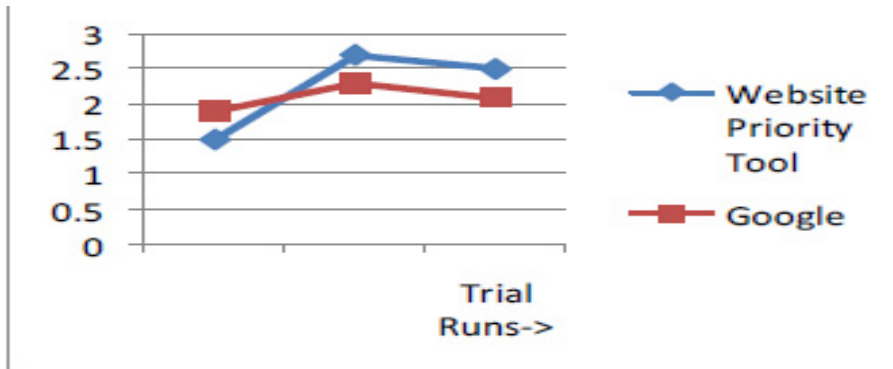

Fig.4. Precision comparison of WPT and Google for E commerce query "Purchase of Apple I Phone 6 Plus"

6. Conclusion and Future Work

This research work presents a semantic and neural based novel approach for E Commerce websites priority ranking with respect to specific product search query. The ranking provided by website priority tool may also help the website designer to optimize the structure of his/her company website by knowing the competitive analysis of his/her website rank. Optimized websites can better serve to society by assisting the users while carrying out online transactions and also helps to increase the revenues of E Commerce Company. The capabilities of proposed SNEC algorithm and Website Priority Tool may further be improved by incorporating various tabs on tool interface such as page loading speed comparison, ease of navigation comparison, online/offline comparison, security comparison etc so that user or E Commerce website owner may specify the criteria to compare the competitive E Commerce websites and to easily determine their rank as per requirement. Moreover incorporation of cloud computing framework such as Hadoop Distributed File Systems (HDFS) may be quite helpful to easily mine the Big Data produced by E Commerce companies.

References

1. ZhuoFan Yang, Yong Shi, Bo Wang, Hong Yan. Website quality and profitability evaluation in ecommerce firms using two- stage DEA model. In proceedings of 1st international conference on data sciences, ICDS2014, Elsevier, pp. 4 – 13.
2. OguzMustapasa, DilekKarahoca, AdemKarahoca, AhmetYucel, HuseyinUzunboylu. Implementation of semantic web on e-learning. In proceedings of WCES 2010, Elsevier, pp.5820-5823
3. WeigangZuo, QingyiHua, WeigangZuo. The application of web data mining in the electronic commerce.In proceedings of IEEE fifth international conference on intelligent computation technology and automation, 2012, pp. 337-339, DOI 10.1109/ICICTA.2012.90.
4. Yanduo Zhao. The review of web mining in e-commerce.In proceedings of IEEE international conference on computational and information sciences, 2013, pp 571-574, DOI 10.1109/ICCIS.2013.158.
5. Shenzihao,Wanghui. Research on e-commerce application based on web mining. In proceedings of IEEE international conference on intelligent computing and cognitive informatics, 2010, pp. 337-340, DOI 10.1109/ICICCI.2010.89
6. Zhiwu Liu, Li Wang. Study of data mining technology used for e-commerce. In proceedings of IEEE third international conference on intelligent networks and intelligent systems, 2010, pp. 509-512, DOI 10.1109/ICINIS.2010.61.
7. G.Lidan Shou, He Bai, Ke Chan, Gang Chen. Supporting privacy protection in personalized web search. IEEE transactions on knowledge and data engineering, Vol. 26, No 2, February 2014, pp. 453-467.
8. Dheeraj Malhotra, Neha Verma. An ingenious pattern matching approach to ameliorate web page rank. international journal of computer applications (0975-8887) New York USA, Vol. 65, No. 24, 2013, pp. 33-39.
9. Shuo Wang, KaiyingXu, Yong Zhang, Feiil. Search engine optimization based on algorithm of bp neural networks, In proceedings of IEEE international conference on computational intelligence and security, IEEE computer society, 2011, pp. 390-394.

10. LI Yaolin, ZHONG Yanhua, NIE Shuzhi. Web user access mode mining based on quantum self organizing neural network. In proceedings of, IEEE international conference on intelligent computation technology and automation, IEEE Computer Society, 2012, pp. 382-385.

11. Erick Gomez –Nieto, Frizzi San Roman, Similarity preserving snippet based visualization of web search results. IEEE transactions on visualization and computer graphics, 2013,pp. 1-14.

12. Yuki Yasuda, Naoto Mukai, Naohiro Ishii. Visualization of page rank algorithm by using multi agent model for education. In proceedings of international conference on advanced applied informatics, IEEE-CPS,2013, pp. 409-410.

13. Meng Cui, Songyun Hu, Search engine optimization research for website promotion.In proceedings of IEEE international conference on IT, CE and management sciences, IEEE computer society, 2011, pp. 100-103.

14. Debajyoti, Manoj Sharma, Ganjman Joshi. Experience of developing a meta semantic search engine,In proceedings of international conference on cloud & ubiquitous computing & emerging technologies, IEEE- CPS, 2013, pp. 167-171.

15. Kun Lin. Applications of web data mining based on the neural network algorithms in e commerce, In proceedings of IEEE international conference on e business learning, IEEE Computer Society, 2010, pp. 509-512.