An Improved Scheme for Organizing E-Commerce-Based Websites Using Semantic Web Mining

Dr S. Vinoth Kumar [1]✉

Email: profsvinoth@gmail.com

Dr H. Shaheen [2]

Email: shaheen66@gmail.com

Dr T. Sreenivasulu [2]

Email: sreesmart@hotmail.com

[1] Department of Computer Science and Engineering, SNS College of Technology, Coimbatore, India

[2] Department of Computer Science and Engineering, St. Peters Engineering College, Hyderabad, India

Abstract

In the running of the Internet world, E- commerce industry has its own benchmark in terms of its rapid growth and has made itself an established sector that is indispensable for every industry to trade and do transactions online. As the world is rushing in a rapid manner, India is slogging in the improvisation of the online market, leading to the lack of customized needs of the customers. Bigger companies are trying to put in a different strategic approach taking that into consideration an approach of blended e-mining along with e-commerce has been devised. It would be a design of the semantic- and neural-based page ranking algorithm [2]. This tool upon launching would be a well-defined approach for e-commerce website ranking [1]. It would also facilitate the customers to find the relevant websites on the top of the page during their search for any particular product or

business. It would be further customized with all the relevant comparison of the other websites in terms of the product quality and price.

AQ1

Keywords

Neural-based page ranking
Website ranking
E-mining

1. Introduction

One of the fastest growing businesses for the past decade is e-commerce [1]. The customer's needs have taken up the next level for satisfying the demands comparing the competitors and to fetch revenue to the company. Researchers made an analysis that in India is also rapidly growing among other countries. The reason behind this sudden growth is due to the increase in the awareness of the Internet-based applications, computer literacy, change of lifestyle, and high rated income [6]. Apart from that are the policies from the companies like trial and exchange feedback of the product, cash upon delivery, and reviews about the product. One of the associate research areas is intelligent neural web based mining would be taking up these features of adaptability and would be able to transit the same to practical application stage where the errors would be minimized to a larger extent and help the customer to extract useful and abstract information from his surfing web pattern [3]. These patterns would help us to do the necessary improvements on the website structure enhancement of the website content making it user-friendly. This paper is going to describe the web mining process deployment to get the maximum benefit of the e-commerce area. It may not only be a user-friendly approach for the customer but also for the data analysts to take a decision on the organizational needs. We have divided the paper into various categories. The phase 2 category is task related. Phase 3 is the identification of the query and analysis. Phase 4 illustrates the objectives of the analysis. Phase 5 details about the analysis method and innovations taken for the page ranking algorithm and website listing tool along with the graphical representation of data [4].

AQ2

AQ3

2. Task Related

The semantic- and neural-based web mining technology leads to a better sophisticated ranking of the e-commerce websites. The main objective for taking up such a blend is that it might be able to assist the customer as well as the organizations to take a more clear approach in terms of the transaction as well as on the data-oriented decision-making. Since the old data mining area was not that successful it has ended up in inventing new areas for progressing further [9]. The intention is to perform a discussion of the quality of the e-commerce websites using the data envelopment analysis model (DEA).

Eventually, they also compared the same with the other models like CCR, BCC, and KH. A large amount of data was taken and was automated for obtaining results on e-commerce websites. Various ideas regarding applications are implied in e-commerce [7]. It also emphasized on the handling of customer behavior pattern and feedback reciprocation. This contributed largely on the optimization of the websites. While handling a large volume of data, most of the important information was hidden during retrieving the data. This hidden information would be of great use for the structuring of the web page and to enhance a better ranking [4]. They also added further that the constant watch of the users, their methodology of watch search patterns, etc. would also be helpful for the website optimization. The primary objective of this research is to improvise the search engine algorithm and to minimize the complexities faced by the user. The semantic and neural network methodology is used for an unbiased ranking model of the e-commerce websites. This model has a wider perspective toward the e-commerce industry through an easier way of user navigation and retrieval of specific information [8].

AQ4

AQ5

3. Identification of the Query and Analysis

The rapid growth of the e-commerce industry remains untold [5]. Every customer is idealized to use only a search engine instead of the web catalog. Search engine may not be able to fetch the exact requirement, as it is a syntactic-based query. It might match and fetch the results based on the frequency count and proximity fetching the data based on search query and web page. This syntactic match may lack semantics producing wrong results

which might fetch either number of unwanted pages or no result at all. Apart from this, the result-producing pages are powered by the search engines which make a very good revenue on the companies listing irrespective of their content, reliability, and relevance to the customer. Few e-commerce companies may not have got the authorization to sell the product. But even then they would have published the same on their websites leading to a confused state of mind for the customer to suffer from not knowing the product in detail along with the other details of warranty, replacement, etc. The reason behind is that the search engines failed to design their structure with reference to the customer's queries and intention of the customer. The other reason is the backpropagation of the errors and retrieval algorithms lead to the biased ranking leading to only the top-based rankings popular.

4. Objectives of the Analysis

The whole aim of the analysis is to enhance the e-commerce website to be ranked in a more better and efficient way using the ranking algorithm through the SNEC process to assist the customer while carrying out online transactions in a more authentic and rational manner. The research implies a semantic- and neural-based approach to deal with the ranking problems. This would optimize the use of web dictionary backpropagation and unbiased ranking process.

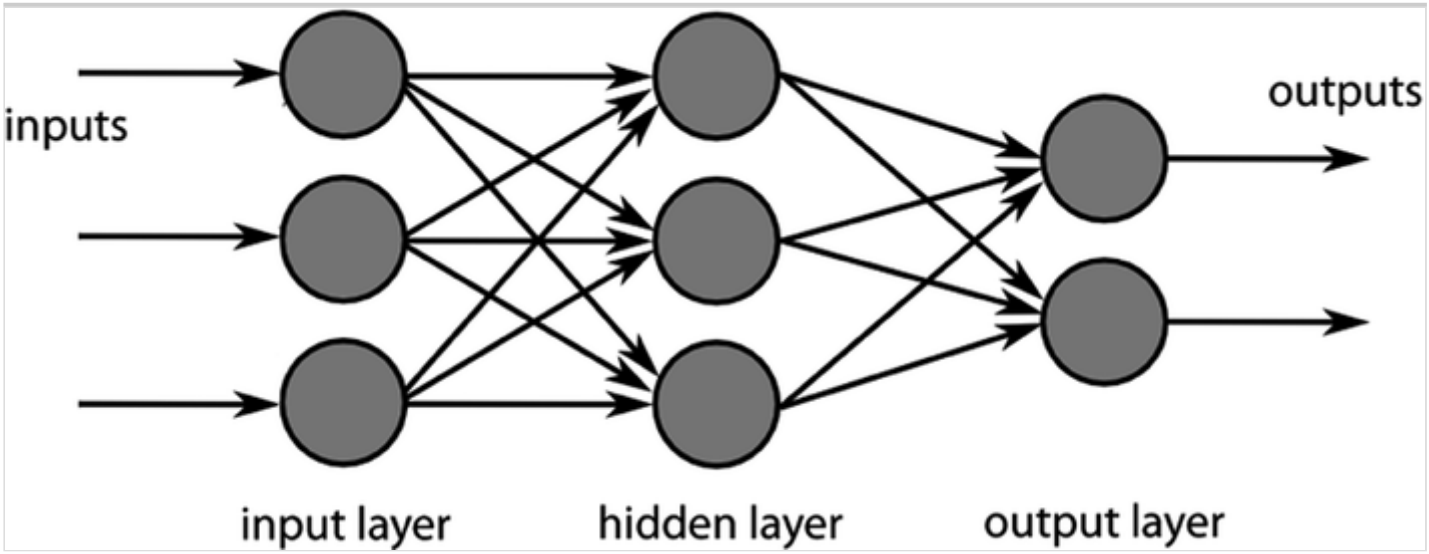
4.1. Analysis Method and Innovations

The research comprises backprocessing the retrieved company information using a profiling and dictionary implementation module to improvise the incomplete entries and data cleaning. The dictionary and the candidate web page are then analyzed by another module called as the content priority module. The primary objective of this module is to check for relevance and to remove unwanted data. Then the query is passed on to the priority module which would check for the priority of the web page based on the customers search and also on the previous searches for the same product by other people. Taking these data the web page is now sent to the next phase of the module called as the semantic module which identifies the user session from other external sources using its algorithm and determines the search to avoid wrong interpretation. One of the most popular NN algorithms is backpropagation algorithm. The BB routine can be fragmented into four key phases followed by which loads of the networks are chosen arbitrarily and the back broadcast routine is employed for estimating the needed modifications. The routine could be disintegrated into the stated four phases as feed onward estimation, back

broadcast for the result layers, back broadcast for the concealed layer, and load-based revision. The scheme is halted on experiencing a fault function which is suitably small. It is very irregular and the basic formula for the back broadcast holds some dissimilarity designed by others but it must be precise and much easier for usage. The concluding phase is load revision which occurs for the overall scheme (Fig. 1).

Fig. 1

Analysis method



AQ6

AQ7

4.2. Nomenclature

Semantic- and Neural-based Algorithm

$\mathbf{X_i}$	User search product.
Min	Minimum length of the keyword X_i.
Max	Maximum length of the keyword X_i.

Y Keyword search specific.

ST The web-based e-commerce document to be scanned.

PD Dictionary with reference to the *T*th document.

TXT Words of the document.

A1 Time spent of the browsing by the other visitors.

A2 Time spent on the web page creation.

F Frequency of the number of keywords found.

NF Not found keywords.

tanθ Linear activation function for the training of the neural network.

MT Mass of the input.

4.3. Module

Module 1:

Step 1: Input from the user.

Step 2: Filtering of unwanted terms from the user.

Step 3: Track the movement of the pattern sequence of the user data.

Step 4: Track the web pages through the search engine.

Step 5: Divide the strings into various words like: Y_1, Y_2, …, Y_n.

Step 6: Determination of minimum and the maximum length of the words

Min := StrLen ($Y1$), Max : = StrLen ($Y1$)

For $k = 1$ to n do

Initialize $F := 0$ and NF:= 0

If ST found in PD then

$F: = F + 1$

Else $F: = \mathrm{NF} + 1$.

Step 11: Remove those web pages where NF > F.

Module 2

Step 12: To evaluate the timestamp $A2$ for the creation of web page.

Step 13: On the beginning of the user session, determine $a1$ which is session duration of current page and determine new value of $A1$ as follows:

If $A1 = 0$ then $A1 = a1$

Else $\mathrm{A1} = (A1 + a1)/2$.

Step 14: Assign a higher priority to web page if $A2$ is low and $A1$ is high.

Step 15: Update the time database of tool with keywords, page address, and $A1$.

Module 3

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

LCS $[c, d] = 0$ if $c = 0$, or $d = 0$ OR

LCS $[c, d] =$ LCS $[c - 1, d - 1] + 1$, if $c, d <> 0$ and $X1c = X2d$ OR

LCS $[c, d] = \max(\text{LCS}[c - 1, d], \text{LCS}[c, d - 1]$, if $c, d > 0$ and $X1c <> X2d$.

Step 17: Class generation using Web Ontology Language.

Module 4

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

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

$\{O\} = \tan \theta \{I\}$.

Step 20: Use sigmoidal function for output evaluation in hidden and output layers as $\{O\} = [(1/1 + e - 1)]$ and summation function as $\sum(C1\text{MT}1 + C2\text{MT}2 + C3\text{MT}3 + C4\text{MT}4 + C5\text{MT}5 + B)$.

Step 21: The error rate is determined by adjusting the weight age of the synapses.

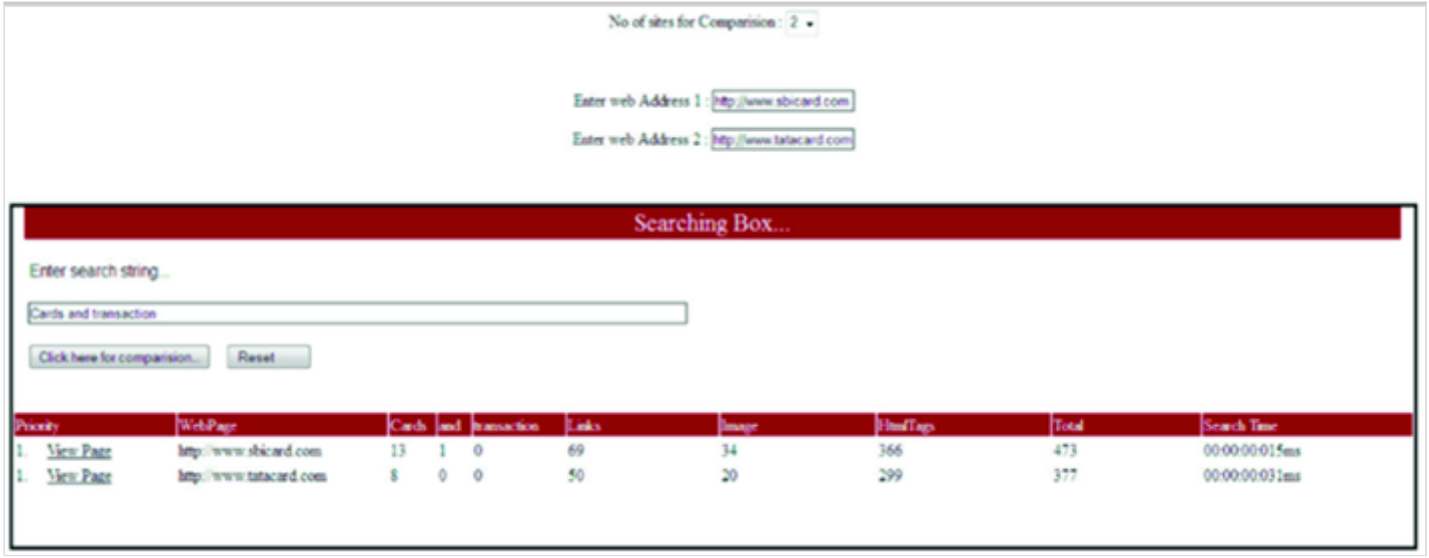
Step 22: At last the web pages are displayed in a decreasing manner in order of the ranking priority.

4.4. Priority on Time Spent

SBPP algorithm mentioned in this exploration research implies the usage of the importance to website priority under ASP.NET framework. Through this research, we would be able to explore more than five e-commerce websites using the design done. The tool will allow the number of entries based on the design. After entering the data the tool would search in accordance with the content and the statistical data (like the number of times the page has been visited, the product specification, etc.) (Fig. 2).

Fig. 2

Priority on time spent



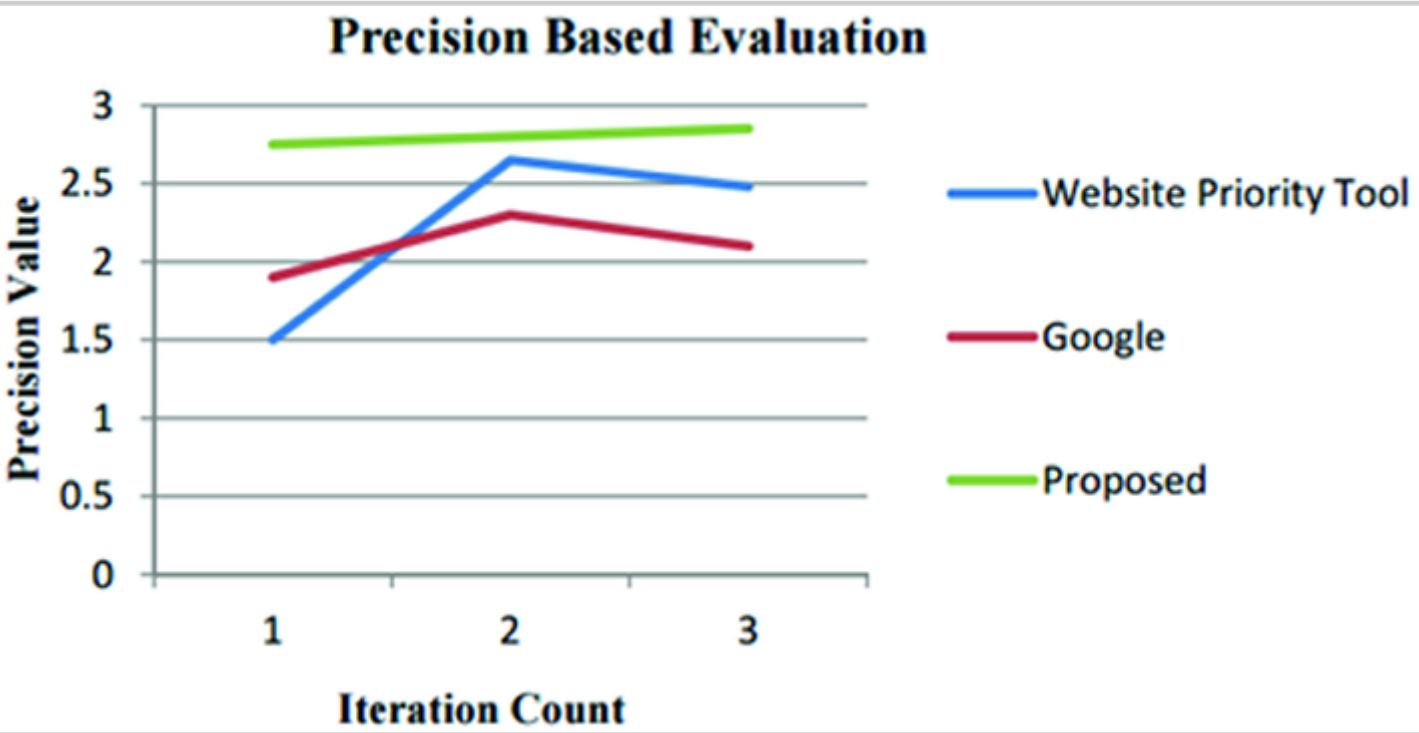
4.5. A Statistical Approach Toward the Research

This segment shows the comparative outcomes of parameters precision and efficiency, respectively. Outcomes were obtained while acting of weighted page rank algorithm on the dataset of Internet pages. Diverse iterations have been completed to test the consistency of the version. Precision: it approaches how properly the website precedence tool (WPT) is working. Internet site precedence device allows evaluation of websites, the use of drop-down container, and seek box to specify a string of unique product. The drop-down field provides as many URL's (uniform resource locator) of the website and after contrast, WPT tool assigns precedence to every candidate website based on the calculation of content priority module, time spent priority module, advice module, and neural priority module. Subsequently, precision is used to measure the consistency of the consequences for each and every time the system runs. Greater the relevancy of the fetched web pages better might be the consistency of the machine. The better consistency of the results implies that the website priority tool is operating correctly. As an

end result, better accuracy of net site priority tool leads to higher precision. Relevancy is calculated with the aid of measuring the space of the statistics. Statistics has been stored in array/matrix shape. Distance will be calculated for every row by comparing it with all different rows. For each row, lesser the gap among rows more relevant will be records and vice versa. Precision values of the proposed machine had been received by way of making use of more than one testing rounds (iterations) about 25 on the statistics set (Fig. 3).

Fig. 3

Precision-based evaluation



Precision primarily based evaluation of the designed version. The foreseen graph represents the correct values for Internet site precedence tool, Google, and the proposed WPR. The line graph here truly shows that the proposed weighted page rank set of rules has high precision values for all the iterations. The graphical layout of both the model's page rank and stepped forward weighted page rank showcases the comparative evaluation which has been

evaluated on the premise of the precision of the simulated outcomes. The upgrades are stated which validates the better conduct of the designed WPR scheme than the prevailing schemes.

5. Conclusion and Future Work

We see a developing interest in the use of the semantic and neural community for solving net programs inside the approaching years. The inherent capability of neuro-semantic techniques in managing indistinct, big-scale, and unstructured information is a great fit for Internet-related problems. Earlier studies have a tendency to consciousness on a way to extract wanted facts from unstructured net facts. Recently, we have seen the use of neural methodologies in building a based web [10]. The semantic Internet is one instance of such. The perception of a structured web can be made extra practical while the concept is employed due to the fact that net records have a tendency to be unpredictable in nature. We count on to peer an integration of gentle computing techniques in semantic Internet methodologies within the near destiny. A genetic set of rules for net software ought to also turn out to be extra famous as Internet applications get large in scale.

References

1. N. Verma, Improved web mining for e-commerce website restructuring, in *IEEE International Conference Computational Intelligence & Communication Technology*, UP, India (2015), pp. 155–160

2. N. Verma, D. Malhotra, M. Malhotra, J. Singh, E-commerce website ranking using semantic web mining and neural computing, in *Proceedings of International Conference on Advanced Computing Technologies and Applications*, Mumbai, India (2015). Procedia Comput. Sci., Science Direct. Elsevier, **45**, 42–51. https://doi.org/10.1016/j.procs.2015.03.080

3. D. Malhotra, N. Verma, An ingenious pattern matching approach to ameliorate web page rank. Int. J. Comput. Appl. **65**(24), 33–39 (2013). https://doi.org/10.5120/11235-6543

4. D. Malhotra, Intelligent web mining to ameliorate web page rank using back propagation neural network, in

Proceedings of 5th International Conference, Confluence: The Next Generation Information Technology Summit, Noida, India (IEEE, 2014), pp 77–81. https://doi.org/10.1109/confluence.2014.6949254

5. S. Sharma, P. Sharma, Use of data mining in various fields. IOSR J. Comput. Eng. IOSR-JCE **16**(3) (2014)

6. N. Padhy, P. Mishra, R. Panigrahi, The survey of data mining applications and feature scope. Int. J. Comput. Sci. Eng. Inf. Technol. IJCSET **2**(3) (2012)

7. H. Kaur, A review of applications of data mining in the field of education. Int. J. Adv. Res. Comput. Commun. Eng. **4**(4), 2319–5940 (2015)

8. S. Bagga, G.N. Singh, Applications of data mining. Int. J. Sci. Emerg. Technol. Latest Trends (2012)

9. N. Jain, V. Srivastava, Data mining techniques. Int. J. Res. Eng. Technol. eISSN: 2319-1163 | pISSN: 2321-7308

10. D. Kumar, D. Bhardwaj, Rise of data mining: current and future application areas. Int. J. Comput. Sci. Issues **8**(5) (2011). ISSN (Online): 1694-0814