Journal of Applied Sciences

ISSN 1812-5654

Journal of Applied Sciences 13 (17): 3504-3509, 2013
ISSN 1812-5654 / DOI: 10.3923/jas.2013.3504.3509

Optimization Model and Algorithm for Web Page Linkage Structure under Electronic Commerce

Zhaoji Yu and Wen Lu
School of Management, Shenyang University of Technology, Shenyang, 110870, China

Abstract: In e-commerce web site, the webpage linkage structure directly influenced the efficiency of obtaining commodity information and the degree of customer's satisfactory. The webpage linkage structure also influenced the online reputation. The aim of the study on the optimization of webpage linkage structure in e-commerce is to solve the above problems. An optimization model of the study for web page linkage structure was established. The optimization objective includes minimizing average searching depth of web linkage structure, maximizing average associating degree among linkage web pages and maximizing significant degree of web pages. To solve above model easily, the transforming approach of the optimization model for web page linkage structure is studied. For solving the optimization model, antibody representing, affinity degree calculating, clone selecting were studied and the solving steps of artificial immune algorithm were designed. The artificial immune algorithm provided a wider leeway of decision-making for a variety of optimization problem. The result showed that, under the condition that the solving precisions by two algorithms were almost the same, the optimal solutions by artificial immune algorithm appeared obvious diversity, by which multi typical schemes can be provided for linkage structure designing in e-commerce web site.

Key words: Electronic commerce, artificial immune algorithm, webpage linkage, structure optimization

INTRODUCTION

With the rapid development of the internet, especially in recent years the mobile phone has gradually popularized on the use of surfing and the E-commerce has presented a new model. However, the structure of web hyperlink in the E-commerce website has directly influenced the speed of implementation for the E-commerce and the efficiency which the customer got information of the product. And it also influenced the satisfaction and reputation of the E-commerce website, especially the management benefit for the E-commerce website. As a trading platform its own design may have impact on online reputation. But the application of optimization model with the web hyperlink structure plays a screening role on linking to website.

MATERIALS AND METHODS

At present, the domestic and foreign scholars have done some research on the design and optimization of the website link structure in the E-commerce (Fang and Holsapple, 2007). The study has done the research that the link structure about website navigation influenced on the application of website. Their result showed that the application-oriented is better than the subject-oriented in the structure of linkage about website navigation. Nakayama *et al*. (2000) proposed a method based on the difference between the website designer and user's accessing behavior. The method can optimize the design of website structure and improve the link structure of the website. Fu *et al*. (2002) proposed a method to recombinant website, but the method only considered the local optimization of website structure. In the literature, the authors found the difference the correlation degree of user access interest and site topology. The authors increased the direct link to the website page in order to realize the optimization of website structure. Wang and Wang (2003) designed path tree spanning algorithm to calculate the page accessibility. A mathematical model was presented to maximize the covariance of visiting rate and accessibility of Web pages. As solving method, path tree spanning algorithm was embedded in the Tabu Search for optimal method. In the literature, Feng and Tao (2007) proposed an optimization model of website structure in the E-commerce based on the planning of 0-1. The model took the maximization of the access frequency as optimization objective based on the basic link structure of website without destroying. Furthermore, it took the number of website link and the shortest path length from home page to each page as the constrain condition. At last, the model given the solution algorithm based on the

Corresponding Author: Zhaoji Yu, School of Management, Shenyang University of Technology, Shenyang, 110870, China

idea of genetic algorithm and particle swarm optimization. Therefore, the classical evolutionary algorithm will lose the ability to adapt the environmental change in the later of evolution, which are the main challenges faced by the evolutionary algorithm in dynamic environment. The artificial immune algorithm is proposed in recent years.

DESCRIPTION TO THE OPTIMIZATION PROBLEM OF THE LINK STRUCTURE IN E-COMMERCE WEBSITE

The optimization problem of the link structure in the E-commerce website needs to consider the following constraints:

- It limits priority level and subordination in the level of web link
- It accords with the relevance of web content. The correlation degree of every direct link webpage is more than a certain threshold value
- It accords with the important constraints in the E-commerce webpage. The mean of importance in the web node is more than a certain threshold value
- Avoid the user's information overload. It refers to limit the number of webpage links. Because the webpage that has many links can make user's information overload. Of course, the different webpage have the different number of links
- Access path depth. It refers to find the shortest path sub graph in the website structure. It contains that the shortest distance from the home page to every webpage. And it shortens the depth of the structure based on the actual demand that the user access the website. The aim is to reduce the path length which the user access a webpage
- Each webpage node has at least one path to link to the home page
- Limit the number of the webpage node that link to the each webpage node. The aim is to prevent the excessive number of the multiple links between two webpage nodes

ESTABLISHMENT THE OPTIMIZATION MODEL OF THE LINK STRUCTURE IN THE E-COMMERCE WEBSITE

To set the variables and symbols

$$x_{ij} = \begin{cases} 1 & \text{a direct link between the webpage node of i and the webpage node of j} \\ 0 & \text{other} \end{cases}$$

r_{ij} = Correlation degree between the webpage of I and the webpage of j
r_0 = Threhold value of the correlation degree in the webpage node
e_{ij} = Derected arc between the webpage of i and the webpage of j
E_L = Set of directed arc with the typical cub ordination in the webpage node
S_i = Importance degree of the mean with the importance degree in the webpage node before the 3 levels
S_0 = Threshold value of the mean with the importance degree in the webpage node before the 3 levels
l_i = No. of layers with the webpage node in the degraph of network structure
l_0 = Threshold value of the layers with the webpage node in the digraph of structure. It means that the shortest distance from home page to every page
L_i = Set of the webpage link node in the i-th layer
g_0 = Threshold value of the webpage node which chain out the webpage node
c_0 = Threshold value of the webpage node which chain into the webpage node
N = Total number of the webpage node in the website
N_i = Total number of the webpage node in the lin layer of i

Establishment of the optimization model in the website link structure

$$\text{Min}\, f_1 = \sum_{i=1}^{N} l_i / N \tag{1}$$

$$\text{Max}\, f_2 = \sum_{i=1}^{N}\sum_{j=1}^{N} x_{ij} r_{ij} \tag{2}$$

$$\text{Max}\, f_3 = \sum_{i=2}^{N} x_{1i} s_i \tag{3}$$

$$l_i < l_j, \quad e_{ij} \in E_L \tag{4}$$

$$r_{ij} \ge r_0 - M(1 - x_{ij}),\ M = +\infty \tag{5}$$

$$\sum_{l_i=1}^{3} s_i / \sum_{i=1}^{3} N_i \ge s_0 \tag{6}$$

$$\sum_{j=1}^{N} x_{ij} \le g_0,\ i = 1, 2, \cdots, N \tag{7}$$

$$l_i \le l_0 \tag{8}$$

$$\sum_{i \in L_{j-1}} x_{ij} \ge 1,\ i = 1, 2, \cdots, N \tag{9}$$

$$\sum_{i=1}^{N} x_{ij} \le c_0,\ j = 1, 2, \cdots, N \tag{10}$$

$$x_{ij} = 0 \text{ or } 1 \tag{11}$$

Conversion of optimization model in the webpage link: The above model is an optimization problem of multi-objective. At the same time, the algorithm coding solve the optimization under some constraints is more difficult. Therefore, Wang *et al*. (2007) think it is necessary to transform the above optimization problem in order to facilitate solving.

It applies the multiplication-division method to integrate the objective function. The Eq. 12 shows that:

$$\operatorname{Min} f = (\sum_{i=1}^{N} l_i / N) / (\sum_{i=1}^{N}\sum_{j=1}^{N} x_{ij} r_{ij} \sum_{i=2}^{N} x_{1i} s_i) \tag{12}$$

Considering the coding have difficulties in the two constraints under the equation of 4, 7 and 10, the study adds the two constraints to the objective function. Therefore, the original optimization model of webpage link structure can transform into the following:

$$\begin{aligned} \min f &= (\sum_{i=1}^{N} l_i / N) / (\sum_{i=1}^{N}\sum_{j=1}^{N} x_{ij} r_{ij} \sum_{i=2}^{N} x_{1i} s_i) \\ &+ M\{\prod_{e_{ij} \in E_L} [\operatorname{sgn}(l_i - l_j) + 1]\} \\ &+ M\{\prod_{i=1}^{N} [\operatorname{sgn}(\sum_{j=1}^{N} x_{ij} - g_0 - 1) + 1]\} \end{aligned} \tag{13}$$

$$r_{ij} \geq r_0 - M(1 - x_{ij}) \tag{14}$$

$$\sum_{l_i=1}^{3} s_i / \sum_{i=1}^{3} N_i \geq s_0 \tag{15}$$

$$l_i \leq l_0 \tag{16}$$

$$\sum_{i \in L_{j-1}} x_{ij} \geq 1, j = 1, 2, \cdots, N \tag{17}$$

$$x_{ij} = 0 \text{ or } 1 \tag{18}$$

$$\operatorname{sgn}(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0, \\ -1 & x < 0 \end{cases}$$

among them, $M = +\infty$

SOLUTION ALGORITHM OF OPTIMIZATION MODEL WITH THE WEBPAGE LINK

Expression of the antibody: The antibody is applied by the two-dimensional coding shows as the Eq. 19. In this equation, the node numbering of the corresponding position and layer in the webpage link structure is expressed as $x_j^{(l)}$.

$$\begin{bmatrix} x_1^{(2)} & x_2^{(2)} & \cdots & x_n^{(2)} \\ x_1^{(3)} & x_2^{(3)} & \cdots & x_n^{(3)} \\ \cdots & \cdots & \cdots & \cdots \\ x_1^{(m)} & x_2^{(m)} & \cdots & x_n^{(m)} \end{bmatrix} \tag{19}$$

In the equaton of 19, $m = l_0 - 1$. It is an assumption that the webpage link structure is basically pyramid shape in order to determine n. That is the number of the webpage node in the layer of l+1 is more than the number of the webpage node in the layer of l. The average number of the webpage node in each layer in addition to the first layer. In this study, changeable range that the number of the webpage node in each layer is expressed as:

$$\Delta = \frac{N-1}{2(l_0 - 1)}$$

The numbering of the homepage node is 1 in the first layer. The maximum number of the webpage node is g_0 and the minimum number of the webpage node is max [2, int (g_0-Δ)] in the second layer considering the threshold value of the webpage node is g_0. In this study, int () represents the integral function. The number of the webpage node is N_3 in the third layer. Therefore, the study can calculate the approximate value of the median with N_3 based on the Eq. 20:

$$N_3 + (N_3 + 1) + \ldots + (N_3 + l_0 - l) = N - 1 - g_0 \tag{20}$$

According to the above formula, the study gets the interval of the number with webpage node in each layer shows as the equation of 21. In this formula, the lower limit to the number of the webpage node in the layer of l is expressed as $\overline{N}_l$. The upper limit to the number of the webpage node is expressed as:

$$\overline{N}_l, [\underline{N}_l, \overline{N}_l] = [N_3 + (l-3) - \Delta, N_3 + (l-3) + \Delta], \quad l = 3, 4, \cdots, l_0 \tag{21}$$

If the total number of the webpage nodes (N) is 28, the threshold value of the number which the webpage node chained out (g_0) was 4, the threshold value of the number to the layers of the webpage node (l_0) was 5 and the threshold value of the number to the webpage link which chained into (c_0) was 3, the study would get $[\underline{N}_2, \overline{N}_2]$ is [2, 4], $[\underline{N}_3, \overline{N}_3]$ is [4, 10], $[\underline{N}_4, \overline{N}_4]$ is [5, 11], $[\underline{N}_5, \overline{N}_5]$ is [6, 12] and the number of the dimension is 36.

If the encoding mechanism was 0-1, the number of the dimensions would be N×N. If it applied the encoding mechanism that this study used for, the number of the dimensions would be $(l_0 - 1)\times c_0\overline{N}_{l_0}$. With the increasing number of the webpage, the encoding mechanism can help to reduce the number of the dimension.

The random generation process of the coding as follows:

- **Step 1:** The number of the random generation with the node in each layer is expressed as $N_l(l = 2,\cdots,l_0)\in[\underline{N}_l,\overline{N}_l]$).

And it meets the condition of:

$$\sum_{l=2}^{l_0} N_l = N-1$$

In addition, the number of the layer with webpage link is 2

- **Step 2:** It can generate the webpage node with the number of N_i+1 and the random position without overlapping with each other in the layer of l by using the random generator if it meets two conditions. The first condition is that the webpage node is not selected and the association degree of the webpage node is more than 0.4. The second condition is that the identifier of the webpage node is zero
- **Step 3:** If $l<l_0, l=l+1$, the process return to Step 2, or to Step 4
- **Step 4**: If the total value of the column was 0, the study would delete the column
- **Step 5:** If the column was the same as the former column, the study would delete the column
- **Step 6:** If the last value of the column was 0 and the former value of the column was the same as the other column, the study would delete the column

For example, the webpage link structure can be seen in Fig. 1. The number in the Fig. 1 represents the number of the webpage node in the link structure of the webpage.

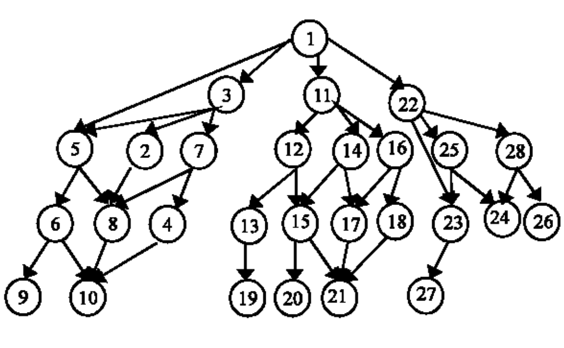

Fig. 1: Link structure of the webpage

CALCULATION OF THE AFFINITY AND THE COLONAL SELECT I

The objective function of optimization model with webpage link is to seek the minimum value. In order to transform the minimum value of objective function into the maximum value, the study defines the affinity function (Aff (Ab$^{(i)}$, Ag)) with the antibody and the antigen as the Eq. 22:

$$m_i = Aff(Ab^{(i)}, Ag) = 1/f \tag{22}$$

the study defines the affinity function (Aff (Ab$^{(i)}$, Ag)) of the antibodies as the Eq. 23:

$$m_{ik} = Aff(Ab^{(i)},Ab^{(k)}) = 1-\frac{\sum_{l=2}^{m}\sum_{j=1}^{n}|x_j^{(l)(i)}-x_j^{(l)(k)}|}{\sum_{l=2}^{m}\sum_{j=1}^{n}|x_j^{(l)(i)}+x_j^{(l)(k)}|} \tag{23}$$

Steve described a basic ideal of clonal selection which adjusted the equations. The adjustment equations showed as the Eq. 24. The number of the antibody is expressed as n. The ratio of the activation is expressed as ω_1. The ratio of the control is expressed as ω_2. The ratio of the death is expressed as ω_3. The density of the antibody is expressed as U_i. The density of the antigen is expressed as V:

$$\frac{dU_i}{dt} = \omega_1 m_i U_i V - \frac{\omega_2}{n}\sum_{j=1}^{n} m_{ij}U_iU_j - \omega_3 U_i \tag{24}$$

The steps of the artificial immune algorithm:

- **Step 1:** According to the above coding method, it generates $(l_0 - 1)\times c_0\overline{N}_{l_0}$ dimensional antibody. The number of the antibody is H. And it forms the population of initial antibody (Ab)
- **Step 2:** According to the equation of 22, the study calculates the affinity (m_i, i = 1, 2, ..., H) of the antibody to the antigen. According to the Eq. 23, the study calculates the affinity (m_{ik}, i, k = 1,2, ..., H) with the antibodies
- **Step 3:** According to the Eq. 24, the study applies clonal selection to each antibody
- **Step 4:** The population of the antibody has regeneration. If the density of the antibody (U_i) was less than the threshold value (y), the antibody should be deleted. The number of deleted antibody is expressed asH_d
- **Step 5:** If the learning process of the antibody met 2 conditions, it would end. The first condition is that the highest density of h antibodies is not less than the

threshold value that is set in advance. The second condition is that the learning error is less than the threshold value of the error or it has reached the maximum number of the iterations. If the learn process did not meet 2 conditions, it would randomly generate H_d antibodies with the method of the Step 1 and repeat the Step 2 to 4

ANALYSIS OF THE ALGORITHM APLICATION

To verify the effectiveness of the above algorithm and the optimization model of the webpage link, the study extracts the typical application example from the optimization practice of the webpage link in the E-commerce website. The importance degree of each webpage node shows in Table 1. The affiliation of each webpage node shows in Table 2. In addition, $g_i = 4$, $c_i = 3$, $i = 1, 2,..., N$; $l_0 = 5$, $s_0 = 0.7$. It should be represented by 0 based on without affect the results.

In order to analyze the result of the artificial immune algorithm, the study applied the methods to randomly run 10 times. the study got the best of 5 solutions to the objective function value showed in Table 3.

The analysis of the Table 3 shows that the mean of the optimal solution in the artificial immune algorithm is basically same as the Genetic Algorithm. The optimal solution of the artificial immune algorithm has more possible state of the node in each layer. In order to compare the diversity of solutions between the artificial immune algorithm and the Genetic Algorithm, the study sets the measurement index of the diversity with the solution shows as the following:

Table 1: Importance degree of each webpage node in the link structure of the webpage

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14
S_i	1	0.9	0.8	0.7	0.8	0.6	0.5	0.4	0.5	0.4	0.8	0.8	0.7	0.6
No.	15	16	17	18	19	20	21	22	23	24	25	26	27	28
S_i	0.5	0.7	0.5	0.4	0.4	0.5	0.5	0.8	0.7	0.8	0.6	0.5	0.6	0.6

S_i: Refers to the importance degree of webpage node, The row of the table represents the number of the webpage node, the column of the table represents the importance degree of webpage node

Table 2: Affiliation degree of each webpage node in the link structure of the webpage

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14
S_i	-	-	-	-	-	-	-	2	-	8	-	-	-	-
No.	15	16	17	18	19	20	21	22	23	24	25	26	27	28
S_i	-	-	11	-	-	-	16	-	-	22	-	-	-	22

S_i: Refers to the affiliation degree of webpage node, The row of the table represents the number of the webpage node, the column of the table represents the affiliation degree of webpage node

Table 3: Optimal solution of the artificial immune algorithm in 5 times random operation

No.	Coding of optimal solution and the function value of the optimal objective																
1	Coding of	2	2	2	2	2	11	11	11	11	11	11	11	22	22	22	22
	optimal	5	4	4	3	3	13	12	12	12	12	12	12	24	23	23	23
	solution	0	8	8	6	7	0	16	16	16	15	15	14	0	27	26	25
		0	10	9	0	0	0	21	19	20	18	17	0	0	0	28	0
	f* =1.1143																
2	Coding of	3	3	3	3	3	16	16	16	16	16	16	16	22	22	22	22
	optimal	8	8	6	2	2	12	12	12	12	12	13	21	23	23	24	25
	solution	10	9	0	5	4	14	11	11	15	15	0	0	26	27	0	0
		0	0	0	7	0	0	17	18	19	20	0	0	28	0	0	0
	f* =1.2152																
3	Coding of	5	5	5	5	0	0	0	16	16	16	16	16	25	25	25	25
	optimal	2	2	9	3	14	14	14	14	15	11	11	11	22	22	23	23
	solution	8	8	4	7	0	12	12	12	21	21	18	17	24	28	28	27
		10	6	6	6	13	13	19	19	19	19	19	20	0	0	0	26
	f* =1.2280																
4	Coding of	2	2	2	2	2	2	14	14	14	14	14	14	14	25	22	22
	optimal	9	9	9	8	5	4	15	12	12	12	12	12	18	27	23	24
	solution	3	6	7	10	0	0	0	11	11	16	16	16	0	0	26	0
		0	0	0	0	0	0	0	17	13	19	20	21	0	0	28	0
	f* =1.3439																
5	Coding of	3	3	3	3	3	16	16	16	16	16	23	23	22	22		
	optimal	2	2	5	4	4	14	14	14	15	11	25	27	24	28		
	solution	8	9	0	7	6	20	20	20	21	21	0	26	0	0		
		10	0	0	0	0	19	18	12	17	13	0	0	0	0		
	f* =1.4092																
All webpage	3, 16, 23, 22, 2, 14, 25, 5, 11																
nodes of each	2, 5, 4, 14, 15, 11, 25, 27, 24, 28, 9, 8, 12, 18, 23, 3, 22, 6, 13, 21																
layer	8, 9, 7, 6, 20, 21, 26, 3, 10, 11, 16, 4, 12, 18, 17, 24, 28, 27, 5, 14, 15, 25																
	10, 19, 18, 12, 17, 13, 20, 21, 28, 6, 26, 7, 9																
	Mean of f = 1.262																

f*: Refers to the optimal solution of the artificial immune algorithm

$$\lambda = \sum_{l=2}^{l_0} (y_l - \bar{N}_l) / \bar{N}_l \quad (25)$$

In this equation of (25), y_i is represented by the total number of webpage node in each layer of l.

The result of the Table 3 shows that the coding of the solution in the artificial immune algorithm presents distinct diversity based on the solving precision with the artificial immune algorithm and the Genetic Algorithm is the basically same. Therefore, it presents the significant differences of link structure in the webpage.

RESULTS AND DISCUSSION

The result of the artificial immune algorithm has a distinct diversity. It is helpful to provide a wider leeway of decision-making for a variety of optimization problem. So the method has been widely used in many areas.

In this study the optimization model of the webpage link structure have considered some optimization objectives. For example, the efficiency of link, the correlation degree of webpage link and the importance of information that the home page directly link the webpage. To some extent, it is helpful to improve the optimization accuracy of the link structure in the E-commerce website. And this study applied the artificial immune algorithm to solve the optimization. It has a prominent characteristic that the solution can implement diversity. The artificial immune algorithm provides more choices to have optimization decision for webpage link structure. The more complex optimization model of the webpage link structure and intelligent optimization algorithm will become research trend in the future in this field. The application of the intelligent optimization algorithm also can popular in other aspects of the E-commerce environment, such as the evaluation of online reputation.

ACKNOWLEDGMENT

This study was supported by Liaoning Provincial Natural Science Foundation of China (201202169) and Program for Liaoning Excellent Talents in University (WR2012003).

REFERENCES

Fang, X. and C.W. Holsapple, 2007. An empirical study of web site navigation structures impacts on web site usability. J. Decis. Support Syst., 5: 476-491.

Feng, J. and H.C. Tao, 2007. An algorithm for optimizing web-site organization based on user access patterns. Microelectron. Comput., 8: 122-124.

Fu, Y., M.Y. Shih, M. Creado and C. Ju, 2002. Reorganizing web sites based on user access patterns. Int. J. Intell. Syst. Account Finance Manage., 11: 39-53.

Nakayama, T., H. Kato and Y. Yamane, 2000. Discovering the gap between Web site designers expectations and users behavior. J. Comput. Networks, 33: 811-822.

Wang, H.F., D.W. Wang and S.X. Yang, 2007. Evolutionary algorithms in dynamic environments. Contr. Deci., 8: 127-131.

Wang, Y.W. and D.W. Wang, 2003. Accessibility and visiting rate based optimal adjustment approach to link structure for e-supermarket website. Inform. Contr., 8: 19-22.