ICEB 2003 Proceedings

International Conference on Electronic Business (ICEB)

Winter 12-9-2003

Recognizing Customer Knowledge Level towards Products for Recommendation in Electronic Commerce

S. Wesley Changchien

Ru-Hui Huang

Recognizing Customer Knowledge Level towards Products for Recommendation in Electronic Commerce

S. Wesley Changchien
Institute of Electronic Commerce
National Chung Hsing University
250 Kuo Kuang Road, Taichung, Taiwan, ROC
swc8@ nchu.edu.tw

Ru-Hui Huang
Department of Information Management
Chaoyang University of Technology
168 Gifeng E. Rd., Wufeng, Taichung County, Taiwan, ROC
s9114605@cyut.edu.tw

Abstract

A powerful online recommendation system in Electronic Commerce (EC) must know its targeted customers well and employ effective marketing strategies. Market research is a very important way to know the customers well. For high-tech products with great variety such as computers, cellular phones, and digital cameras, customers' knowledge level towards products may have a decisive influence on their purchase decision. While many online recommendation systems focus on utilizing data mining techniques in user profile and transaction data, this paper presents a method for recognizing customer knowledge level as a preprocess for more effective online recommendation in EC. The method consists of two Back Propagation Networks (BPN) and predicts based on customer characteristics and online navigation behaviors. A simple simulated digital camera EC store case study was conducted and the good preliminary result implies the good potential of the proposed method.

Key words: Product Knowledge, Web Usage Mining, BPN, EC

1. Introduction

While product recommendation and customization become more and more difficult due to the great and increasing variety of products available on the Internet, e-marketers endeavor to provide effective approaches for product recommendation and personalization. Approaches include: using data mining techniques to conduct market fragmentation and recommend on the basis of the consumer behaviors of each market segment; mining of customers' preference rating for offered products to automatically generate product recommendation through personalized direct marketing; web usage mining of consumer behaviors to provide product recommendation; and the use of segmentation and predictive modeling techniques for database marketing, etc. In the literature, many recommendation systems were built to provide the most appropriate products or services to meet a customer's needs by analyzing his personal preferences based on his profile and purchasing history. However, people with similar profiles may have different preferences, and past purchase patterns may no longer exist. Therefore, it may not be effective to recommend simply according to customer profile and purchase history.

Of various factors affecting customers and marketing effectiveness, customer knowledge towards products, which consists of familiarity and expertise, has a decisive influence on what categories of products are suitable to be recommended to the customer.

Besides, product description and electronic catalog design are also very important in recommendation. The presentation of products for experts or novices must be different. It is unintelligible to use abstruse wording for the novice, and simple description and specification may not be able to satisfy the customers' professional requirements. Therefore, present and recommend products according to customers' knowledge level in EC may better suit their needs and promote the sales.

This paper proposes a method for recognizing customer knowledge level towards products (Figure 1). The method includes two neural networks; the first neural network learns to classify customers by customer profiles, while the second neural network learns to predict the customer's knowledge level towards products. Both neural networks are BPNs. Since the static user profile only provides basic properties of customers, they alone are insufficient for predicting a customer's knowledge level. A specifically designed questionnaire for customer knowledge level towards the products is combined with the user profile as the inputs for the first BPN. In addition, two sets of navigation patterns, mined from the web logs for experts and non-experts, plus the output of the first BPN (customer profile characteristic) are fed as the inputs to the second BPN. The output of the second BPN indicates the customer's knowledge level.

We begin by reviewing previous works related to our research in Section 2. In Section 3, the proposed method for recognizing a customer's knowledge level towards products is presented. The website implemented for an EC store case study is introduced in Session 4. The experiment results are described in Session 5. Finally, we summarize our contribution and give suggestions for future research in Section 6.

2. Background

2.1 Product knowledge

Customers differ in their prior knowledge about products, which consists of familiarity and expertise as defined by Alba and Hutchinson [1], where familiarity refers to the acquaintance level towards products and accumulated purchase experiences, and expertise means the professional know-how and understanding. Furthermore, others researchers also defined different categories of product knowledge including subjective knowledge, objective knowledge, and experience-based knowledge. This paper focuses on the recognition of objective knowledge and experience-based knowledge to enhance personalized marketing effectiveness.

Customers in online environment can be roughly divided into two groups, one is expert group and the other is non-expert group. These two groups differ in the criteria for selecting products and in the product searching behaviors. Alba and Hutchinson [1] and Rao [8] both conducted researches on how product knowledge influences customers on information processing. In EC, customers' knowledge level towards products can potentially be recognized by analyzing customers' profile and navigation patterns.

2.2 Web usage mining

In online environment, the website visitors click hyperlinks that are of interest to them and the access patterns to the website may imply some of the customer behaviors. Web usage mining is one of the data mining techniques for the discovery of behavior patters from web logs. The application of web usage mining is generally divided into two types, personalized recommendation and web site design improvement. Cho et al. [5] have proposed a personalized recommendation methodology based on web usage mining. Their proposed decision tree induction demonstrated good effectiveness and quality of recommendations. Spiliopoulou and Faulstich [13] developed a data mining system for finding the interesting patterns from the web access log. Perkowitz and Etzioni [7] improved a site by creating customized index pages and clustering pages of a site that were frequently visited together in a single session. Simith and Ng [12] evaluated the feasibility of using a self-organizing map (SOM) to mine web log data and provided a visualization tool to assist user in navigation.

Therefore, mining customers' navigation patterns from web usage log may play a decisive role in recognizing customer knowledge level towards products.

2.3 Back Propagation Networks

Human brains learn via a neural network that consists of neurons, axons, dendrites and synapses. The brain receives inputs, analyzes them and outputs feature or pattern recognition information [9]. There are some researches have gone into creating artificial neural networks that learn in similar ways in recent years.

There are several types of artificial neural networks in use today; one of these types, BPN [11] is widely used in prediction and measurement, such as stock price prediction [2], option market trade [4], fraud detection [10], bankruptcy prediction [6], and medical diagnosis [3].

This paper proposes a method which consists of two BPNs and web usage mining to recognize a customer's product knowledge level as will be introduced next.

3. A Proposed Method for Recognition of Customer Knowledge Level towards Products

3.1 The method for recognizing customer knowledge level

In this paper, we propose a method for recognizing customer knowledge level towards products. Figure 1 shows the recognition mechanism using BPN and web usage mining.

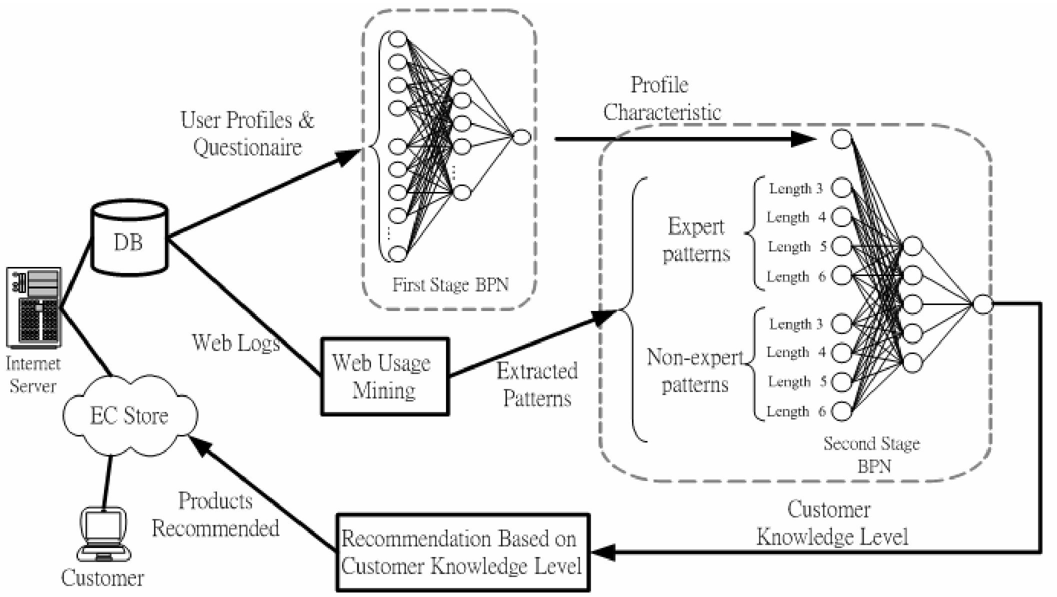

Figure 1. The proposed method of recognizing customer knowledge level using BPN and web usage mining

After consumers sign up, customer profiles and questionnaire results recorded are the inputs of the first neural network which classifies customers. The output value is between 0 and 1, where a value close to 1 means there is a high possibility that the customer has high knowledge level towards products. However, two persons with the same background do not assure that they have the same knowledge level towards a specific type of products. Hence, the output data of the first neural network is only one of the factors considered in recognizing customer knowledge level towards products.

Besides the customer profiles, every time when a customer visits the web site, the system will record all the navigation information in the log file, including the URLs the customer visited, IP addresses, time entering the web page, time leaving the web page, etc. Through sequence pattern mining on the log data, the frequent navigation paths on the web site can be extracted. The frequencies of patterns with lengths 3-6 along with the output of first BPN are the inputs of the second BPN. The output of second BPN indicates the customer's final knowledge level towards products. Once we recognize a customer's knowledge level towards products, the introduction of products, marketing, and recommendation can be personalized according to his knowledge level.

The procedure of recognizing customer knowledge level can be implemented on line in EC. The training data of the BPNs have a critical influence on the performance of recognizing customer knowledge level. The fallowing section explains the procedure of recognizing customer knowledge level with our proposed method.

3.2 The procedure of recognizing a customer's knowledge level

1. Collect training data for the first BPN

Collect the profiles and navigation data of both typical expert and non-expert customers of the web site. The data then are divided into training data and test data sets.

2. Training of the first BPN

The customer profiles and answers to the questionnaires are the inputs of the first BPN for roughly classifying the customers. The only output value is between 0 and 1, where a value close to 1 means the customer may have a high possibility to possess high knowledge level towards products.

3. Classification of the web page contents

EC web sites usually contain dynamic web pages which connect and retrieve data from a database. Web pages retrieving different product details are regarded as different web pages. For example, "pdtl.asp" is a web page to show the details of products, "pdtl.asp?p=A" shows the details of product A, and "pdtl.asp?p=B" presents the details of product B. Although they all belong to the same web page, they should be considered as different web contents since the URLs are different. Consequently, the more products we have, the more web pages will be created. This will lead to a difficult in finding frequent navigation sequence patterns for experts and non-experts. Product classification may be used here to solve this problem. For instance, products may be classified into fewer categories, such as professional, popular, and novice classes, and so on. This will reduce the total number of various navigation sequence patters, and thus increase the support count of popular sequence patterns.

4. Web usage mining

Since the sequences of travel paths on web site are important, a sequence mining method is used to extract frequent navigation sequences of various lengths. With sequence pattern mining, sequences can be drawn specifically for experts and non-experts from the log file, and those sequences which are common between experts and non-experts will not be used since they can not be used to differentiate experts from non-experts.

5. Training of the second BPN

The sequence patterns with lengths 3-6 (3-6 consecutive URLs) for both experts and non-experts plus the profile characteristic are the inputs and the knowledge level, ranged between 0 and 1, is the output for the second BPN.

Following the five steps above, once a new customer registers and leaves some navigation sequences at the web site, the developed method can on line recognize the knowledge level of the new customer towards our products, and then the corresponding recommendation can proceed. However, the recommendation based on customer's knowledge level is not attempted in this paper.

4. A Case Study in Digital Camera EC Store

To testify the effectiveness of the proposed method, a case study in a simulated digital camera EC store is conducted.

4.1 The interface of the EC store

In collecting the training data, selected customers are asked to sign up, fill up the profile, answer the questionnaire which is composed of some simple True/False questions about digital camera, and leave some navigation sequences at the web site.

In the experiment, 30 typical experts and non-experts were selected to visit the developed web site and their profiles, answers to the questionnaire, and navigation logs were recorded. Figure 2 is the interface of the developed simulated EC store.

4.2 Training of the first BPN

For the first BPN, 7 different sets of training and test data were randomly generated for both experts and non-experts, where there are 27 and 3 records for training

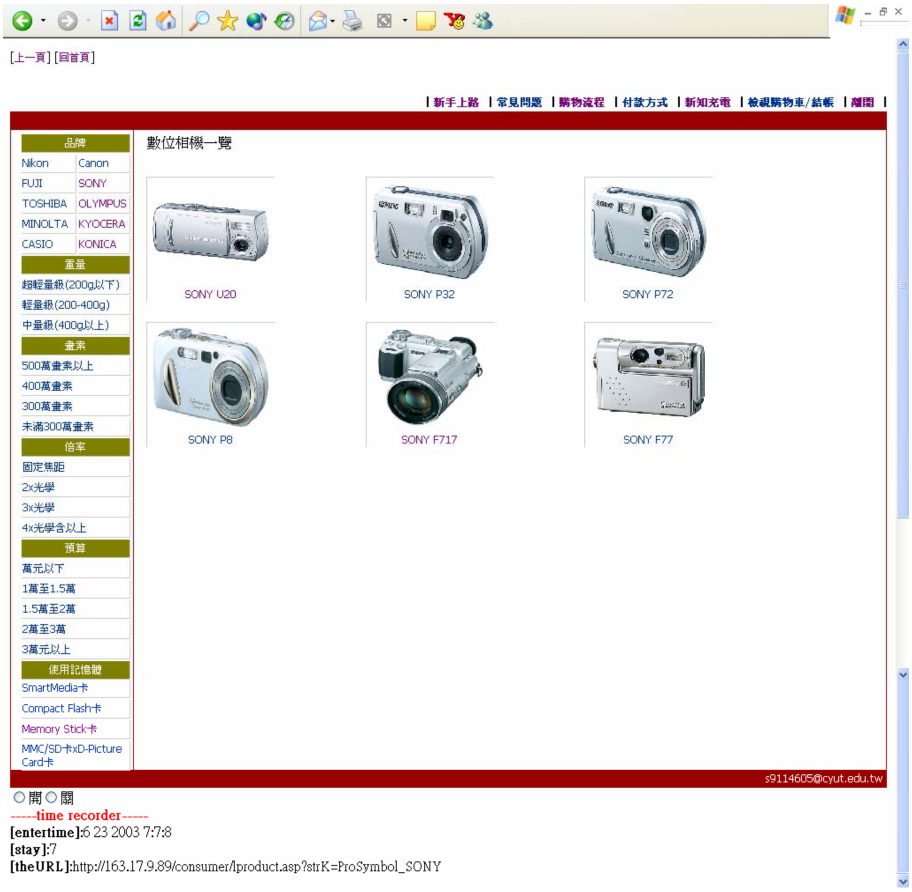

Figure 2. The interface of the simulated EC store

and test purposes for each set of data, respectively. The output of the first BPN denotes the customer profile characteristic, and the more its value is close to 1, the more possibility he has expert knowledge towards the products. The output then becomes one of the input nodes of the second BPN.

4.3 The navigation recorder

To collect customers' navigation data, we developed a navigation recorder which is a plug-in of the web site. While visitors click on the web pages, it records customers' navigation details. The 30 selected experts and non-experts were asked to navigate the EC store, their navigation sequences were recorded for further web usage mining. Table 1 shows some of the navigation details collected. Each record consists of:

1. userID: The user ID.
2. sessionID: While a client connects to the web server, the server will assign a session ID to the client. Though session ID, we can trace the navigation sequences of each session for each customer.
3. theURL: The URLs which a customer has visited.
4. userIP: The IP that each customer used to connect to the web site.
5. TimeIn: The time when the customer enters the web page.
6. TimeOut: The time when the customer leaves the web page.
7. S: Duration of the stay at the web site. Subtract the time entering the web page from the time of leaving the web page and take out the period of time during which the customer's focus is on another URL. By monitoring if a customer's focus is on another application, we can detect if the customer is inattentive while navigating the current URL.

Table 1. Some of the navigation records collected

no	userID	sessionID	theURL	userIP	TimeIn	TimeOut	S
1	taco	1050883611	http://163.17.9.89 /lproduct.asp	211.74.236.25	03/6/3 pm08:51:30	03/6/3 pm08:51:56	25
2	taco	1050883611	http://163.17.9.89 /lproduct.asp?str K=K_2	211.74.236.25	03/6/3 pm08:51:57	03/6/3 pm08:52:23	26
3	taco	1050883611	http://163.17.9.89 /lproduct.asp?str K=P_K	211.74.236.25	03/6/3 pm08:52:24	03/6/3 pm08:52:33	8
4	taco	1050883611	http://163.17.9.89 /logout.asp	211.74.236.25	03/6/3 pm08:52:33	03/6/3 pm08:52:38	4
5	wice	1050883677	http://163.17.9.89 /lproduct.asp	163.17.9.75	03/6/4 pm05:33:51	03/6/4 pm05:34:11	20
6	wice	1050883677	http://163.17.9.89 /lproduct.asp?str K=P_N	163.17.9.75	03/6/4 pm05:34:11	03/6/4 pm05:34:12	1
7	wice	1050883677	http://163.17.9.89 /lproduct.asp?str K=P_C	163.17.9.75	03/6/4 pm05:34:13	03/6/4 pm05:34:17	4
8	wice	1050883677	http://163.17.9.89 /a_new.asp	163.17.9.75	03/6/4 pm05:34:18	03/6/4 pm05:34:25	7

By sequence pattern mining from the customer navigation data which the navigation recorder has recorded, the navigation patterns for both experts and non-experts can be obtained.

4.4 Product classification and sequence pattern coding

Product classification is applied in order to decrease the total number of different navigation sequences. In this case study, we classified products into four groups by product specification. URLs showing the same classified products on the same web page file are assigned with the same web page code. Table 2 shows some of the web page codes according to the product classification and the URL.

Table 2. Web page coding based on product classification and URL

no	file_detail	theUrl	code
1	Overview: products	http://163.17.9.89/consumer/lproduct.asp	00
2	By brand - Nikon	http://163.17.9.89/consumer/lproduct.asp?strK=ProSymbol_Nikon	01
3	By brand - Canon	http://163.17.9.89/consumer/lproduct.asp?strK=ProSymbol_Canon	02
4	By brand - FUJI	http://163.17.9.89/consumer/lproduct.asp?strK=ProSymbol_FUJI	03
5	By brand - SONY	http://163.17.9.89/consumer/lproduct.asp?strK=ProSymbol_SONY	04
6	By brand - TOSHIBA	http://163.17.9.89/consumer/lproduct.asp?strK=ProSymbol_TOSHIBA	05
7	By brand - OLYMPUS	http://163.17.9.89/consumer/lproduct.asp?strK=ProSymbol_OLYMPUS	06
8	By brand - MINOLTA	http://163.17.9.89/consumer/lproduct.asp?strK=ProSymbol_MINOLTA	07
9	By brand - KYOCERA	http://163.17.9.89/consumer/lproduct.asp?strK=ProSymbol_KYOCERA	08
10	By brand - CASIO	http://163.17.9.89/consumer/lproduct.asp?strK=ProSymbol_CASIO	09
11	By brand - KONICA	http://163.17.9.89/consumer/lproduct.asp?strK=ProSymbol_KONICA	10
12	By weight - light	http://163.17.9.89/consumer/lproduct.asp?strK=Kheavy_1	11
13	By weight - common	http://163.17.9.89/consumer/lproduct.asp?strK=Kheavy_2	12
⋮	⋮	⋮	⋮
472	Move product from market basket	http://163.17.9.89/consumer/dbasket.asp?ProNo=13	72
473	Pay up	http://163.17.9.89/consumer/lpay.asp	73
474	Finish pay up	http://163.17.9.89/consumer/lpayOk.asp	74

4.5 Mining the navigation patterns

By mapping the sequence of URLs that a customer visited to the code table (Table 2), a navigation sequence can be converted into a coded sequence such as "26604864006048265256 26," where every two digits denote a URL. With sequence pattern mining, frequent sequence patterns can be drawn specifically from experts and non-experts. Through a sliding window, we can get all the sequence patterns of length 3-6. Table 3 shows the patterns with lengths 3-6 for the specific sequence "2660486400604826525626."

Table 3. An example of obtaining sequence patterns by sliding window

Navigation	**2660486400604826525626**
Length 3	266048 604864 486400 640060 006048 604826 482652 265256 525626
Length 4	26604864 60486400 48640060 64006048 00604826 60482652 48265256 26525626
Length 5	2660486400 6048640060 4864006048 6400604826 0060482652 6048265256 4826525626
Length 6	266048640060 604864006048 486400604826 640060482652 006048265256 604826525626

After obtaining the sequence patterns from experts and non-experts, and removing those patterns which are common between experts and non-experts, expert specific and non-expert specific patterns are found.

4.6 Compute the support count for expert and non-expert specific patterns.

Calculate the support counts of all experts and non-experts specific patterns using the collected training data. Then divide the support counts by the value of the maximal support count resulting in the normalized values. Since there are only hundreds of patterns, they are all adopted in the experiment. For a large number of different patterns, only significant patterns with normalized support count values larger than a pre-defined threshold need to be included in the inputs of the second BPN.

4.7 Training of the second BPN

With sequence pattern mining, sequences can be drawn specifically from experts and non-experts, and those sequences which are common between experts and non-experts will not be used for recognition for the second BPN. With the sequence patterns with lengths 3-6 (3-6 consecutive URLs) for both experts and non-experts plus the profile characteristic as the inputs and the knowledge level (0 or 1) as the output value, we can train the second BPN and use it to predict on line if a new customer has high knowledge level soon after he registers and navigates the EC store. All the 27 experts' or non-experts' data were fed into the second BPN with the output values of 1 and 0, respectively to train the second BPN. The rest of the 3 experts' or non-experts' data were used as the test data for the BPN. The experiment results will be provided in the following section.

5. Experiment results

This case invited some experts who have often posted articles on the digital camera discussion panel and some non-experts who are very unfamiliar with digital cameras.

To avoid bias in separating the data into training and test data sets, this case study divided the data set into two subsets which consists of a training (90%) and a test (10%) data sets. Similarly, totally 7 different ways of dividing the data set were repeated (another 3 inappropriate data sets were disregarded). Table 4 shows the test result of one out of the 7 data sets.

Table 4 The test results for one set of training and test data

	exp15	exp19	exp23	non11	non14	non25
exp	0.98063	0.041385	0.97488	0.44754	0.44716	0.45059
nonexp	0.043285	0.66249	0.054541	0.73465	0.73492	0.73251
	correct	incorrect	correct	correct	correct	correct

Accuracy: 83.33

In the 7 sets of training and test data for the proposed method, the test results show that the average prediction accuracy is 90.48%. Furthermore, we changed the ratio of training data vs. test data to 6:4, and the average accuracy is 83.33 %.

Although the accuracy is highly dependent on the training and test data sets, the good preliminary results show the high potential of the method to be applied in online customer knowledge level recognition, on which further recommendation can be based.

6. Conclusion and Future Works

This paper proposes a method for recognizing customer knowledge level towards products. The method consists of two BPNs, the first one is trained to differentiate customers based on user profile and specifically designed questionnaire and the second one is designed to predict the knowledge level, by taking into account additional navigation sequences.

In the experimentation, we divided the data set into training data and test data with the ratio of 9:1 repeatedly for 7 different sets, the accuracy is 90.48%. The results show the high potential of the method in recognition of customer knowledge level towards products.

Once a new customer registers and leaves some navigation sequences at the web site, the developed method can on line recognize the knowledge level of the new customer towards our products, and thus further online personalized product recommendation can be generated and offered.

The preliminary result of the case study has shown the high potential of the proposed method in EC. As for the future works, interesting directions to improve include the questionnaire design, sensitive analysis of how the input data will affect the prediction accuracy, and application of a large real case study to verity the effectiveness of the proposed approach.

References

[1] Alba, J. W. & Hutchinson, J. W., “Dimensions of consumer expertise,” Journal of Consumer Research, 1987, 13(3), 411-454.

[2] Baba, N. & Kozaki, M., “An intelligent forecasting system of stock price using neural networks,” IJCNN-92, 1992, I, 371-377.

[3] Baxt, W. G., “Use an artificial neural network for data analysis in clinical decision-making: the diagnosis for acute coronary occlusion,” Neural Computation, 1990, 2, 480-489.

[4] Bergerson, K. & Wunsch, D. C., “A commodity trading model based on a neural network-expert system hybrid,” IJCNN-91, 1991, 1, 289-293.

[5] Cho, Y. H., Kim, J. K. & Kim, S. H., “A personalized recommender system based on web usage mining and decision tree induction,” Expert Systems with Applications, 2002, 23, 329-342.

[6] Odom, M. D., “A neural network model for bankruptcy prediction,” IJCNN-89, 1989, 2, 163-168.

[7] Perkowitz, M. & Etzioni, O. ”Towards adaptive websites: conceptual framework and case study,” Artificial Intelligence,2000,118,245-275.

[8] Rao, A.R. & Monroe, K. B., “The moderating effect of prior knowledge on cue utilization in product evaluations,” Journal of Consumer Research, 1988, 15(2), 253-264.

[9] Rao, V. R. & Rao, H. V., C++ Neural networks and fuzzy logic, 1993.

[10] Rochester J. B. & Douglass, D. P., “New business for neurocomputing,” I/S Analyzer, 1990, 28(2), 1-5.

[11] Rumelhart, D. E., Hinton, G. E., & Williams, R. J., “Learning internal representation by error propagation,” in Parallel Distributed Processing, 1986, 1, 318-362, D.E. Rumelhart and J. L. McClelland (Eds.), M.I.T. Press, Cambridge, MA.

[12] Smith, K. A. & Ng, A., “Web page clustering using a self-organizing map of user navigation patterns,” Decision Support Systems, 2003, 35, 245-256.

[13] Spiliopoulou, M. & Faulstich, L.C., “WUM: a tool for web utilization analysis,” Proceedings of the International Workshop on the Web and Databases, Valencia, Spain, 1998, 184-203.