Expert Systems with Applications 30 (2006) 682–693

Expert Systems with Applications

www.elsevier.com/locate/eswa

# Assessing users' product-specific knowledge for personalization in electronic commerce

Shuchih Ernest Chang[a], S. Wesley Changchien[a,*], Ru-Hui Huang[b]

[a]Institute of Electronic Commerce, National Chung Hsing University, 250 Kuo Kuang Road, Taichung City 402, Taiwan, ROC

[b]Department of Information Management, Chaoyang University of Technology, 168 Gifeng E. Rd., Wufeng, Taichung County, Taiwan, ROC

**Abstract**

While many electronic commerce (EC) companies are adopting one-to-one marketing approaches using various personalization technologies to make their products and services unique for the purpose of attracting and retaining customers and improving their completion edges in the EC ecosystem, which, nevertheless, has low entrance barriers for new players to join and further intensify the competition, none or few of them consider a fundamental issue—the user's product-specific knowledge. Our research proposed to add this new domain of the customer's knowledge on appropriate target products into the personalization process as a part of the overall EC strategy for businesses. In this paper, we present our initial design for assessing the user's product-specific knowledge using the proposed innovative method for detecting it directly in a non-intrusive way without asking users to answer or fill out any types of questionnaires. Our method is based on customer's on-line navigation behaviors by analyzing their navigation patterns through pre-trained artificial neural networks. An empirical study designed for a case of EC store selling digital cameras was conducted in our research to prove the concept, and a good preliminary result was derived from the study.

For the purpose of comparing the performances between the conventional approach of using questionnaire and the proposed innovative approach of navigation pattern mining, a questionnaire based approach for evaluating the user's product-specific knowledge was designed and incorporated into our knowledge level assessment system (KLAS). Our study result shows that although the pure questionnaire-based KLAS is intrusive and may not be accepted by some users, for those users willing to complete the questionnaire, the proposed navigation pattern approach can be combined with the questionnaire-based approach to create a hybrid KLAS which has a significantly improved accuracy rate in detecting the customer's product knowledge level.


*Keywords:* Personalization; Electronic commerce; Product knowledge; Web usage mining; Artificial neural network

## 1. Introduction

In recent years, people's life and living styles have been deeply influenced by Internet, which enables electronic commerce (EC) for companies and their business partners to conduct business and perform electronic transactions (Lin & Lu, 2000; Liu & Arnett, 2000). In addition to the purchase of products and services over the Internet, EC also encompasses all electronically conducted business activities, operations, and transaction processing within and cross companies. Through EC, companies can alleviate constraints (upon time, space, and cost) to enhance the way they connect to and interact with their EC counterparties by serving customers and collaborating with business partners electronically and intelligently. To catch the revolutionary opportunity and benefit of EC, an explosive number of companies are competing in the EC ecosystem, which, nevertheless, has low entrance barriers for new players to join and further intensify the competition. Thus, for the purpose of attracting and retaining customers and then improving their completion edges, some EC companies take advantage of differentiation and personalization technologies to make their products and services unique and to tailor their products and services for specific user preferences. For example, through personalization, businesses can research on customer's behaviors for developing appropriate marketing strategies, and then delivering suitable products and services to the targeted customers accordingly. Wind and Rangaswamy (2001) found that the opportunity and capability to offer

* Corresponding author. Tel.: +886 4 22859465; fax: +886 4 22859497.
*E-mail address:* swc8@nchu.edu.tw (S.W. Changchien).

0957-4174/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2005.07.021

consumers a flexible and personalized relationship is probably one of the most important advantages among all possible benefits offered by EC to businesses. It is shown that personalization can ultimately enhance customer's satisfaction level and loyalty, and the increase in each customer's visiting frequency can further create more transaction opportunities and benefit the Internet businesses (Lee, Liu, & Lu, 2002).

From the consumers' point of view, the Internet has become a major channel to the worldwide sources of information. While the Internet traffic has been increasing rapidly since 1997, at the range between 70 and 150% annually (Odlyzko, 2003), it is estimated that the amount of information available from Internet doubled every 18 months, and the number of home pages is even increasing in a faster rate (Yang, Yen, & Chen, 2000). This fact is causing users a serious problem of information overload when they try to retrieve information from the dynamically and continuously growing web resources. Therefore, the need from web users in identifying and using more intelligent systems or tools for conducting information gathering and information filtering from the huge size of web related sources is on the rise (Li & Zhong, 2004). In regard to products and services, different people have different and/or various needs, interests, and preferences; nevertheless, the taste and inclination of a person on products and services may also change or evolve with time. Thus, the 'one-to-one marketing' strategy was proposed to provide personalized service in the EC environment (Allen, Kania, & Yaeckel, 1998; Weng & Liu, 2004). 'If we have two million customers, then we should have two million shops on our website', said by Jeff Bezos, CEO of Amazon, may serve as an example to show the importance and the value of the personalization strategy in EC environment. Personalization technology can give users a better, in terms of efficiency and effectiveness, EC experience since they do not have to browse through all the irrelevant noise.

In general, there are two major approaches to provide personalized information: content-based and collaborative filtering (Aggarwal, Wolf, Wu, & Yu, 1999; Yu, 1999). In the content-based approach, it matches the content of candidate items against the user profile, which is constructed by analyzing the content of items that the user has favored in the past or user's personal information and preferences. Some recommendation systems, which are used by EC companies to suggest products and provide information to customers, operate based on this approach, such as NewsWeeder (Lang, 1995) and Infofinder (Krulwich & Burkey, 1996). In the collaborative filtering approach, it identifies other users that have showed similar preference to the given users and provides what they would like. Several recommendation systems are developed based on this approach, such as Tapestry (Goldberg, Nichols, Oki, & Terry, 1992), GroupLens (Konstan, Miller, Maltz, Herlocker, Gordon, & Riedl, 1997), Ringo (Shardanand & Maes, 1995), PHOAKS (Terveen, Hill, Armento, McDonald, & Creter, 1997), and Siteseer (Rucker & Polenco, 1997). While the content-based personalization suffers limitations in dealing with non-text multimedia resources (such as movies, music, etc.) and in making classified recommendations other than the localized domain specified by a user's profile/preference, the collaborative filtering approach is unable to provide new items to a user and unsuited to a user with changing or evolving preferences. Various renovated or hybrid approaches were proposed to cope with the shortcomings of content-based personalization and collaborative filtering, and to increase the accuracy of recommendation systems, by integrating content-based approach and collaborative filtering approach (Changchien, Lee, & Hsu, 2004; Weng & Liu, 2004), applying data mining techniques (such as association rules mining) to collaborative filtering (Kim, Cho, Kim, Kim, & Suh, 2002; Lee, Kim, & Rhee, 2001; Wang & Shao, 2004; Wang & Thao, 2003), and combining collaborative filtering based on item and collaborative filtering based on user approaches (Li, Lu, & Xuefeng, 2005).

Although all the above-mentioned recommendation systems share the same spirit of assisting in the user's search of items of interest, none of them address a fundamental issue—the user's product-specific knowledge. Our research proposed to add this new domain of customer's knowledge on all potential products into the personalization process as a part of the overall EC strategy for businesses. In this paper we present our initial design for assessing the user's product-specific knowledge. Since a user's product-specific knowledge level on various products varies, we proposed an innovative method for detecting it directly in a non-intrusive way without asking the user to answer or fill out any types of questionnaires. Our method is based on the customer's on-line navigation behaviors by analyzing their navigation patterns through pre-trained artificial neural networks. An empirical study designed for a case of EC store selling digital cameras was conducted in our research to prove the concept, and a good preliminary result was derived from the study. This automatic and non-intrusive approach for evaluating the customer's product-specific knowledge was incorporated into our personalized promotion decision support system (Changchien et al., 2004), which used data mining techniques in accordance with marketing strategies to help the business prepare the highly potential and suitable promotion products for each individual customer.

The subsequent sections of this article are organized as follows. Section 2 describes product knowledge, web usage mining, and back propagation networks. Section 3 proposes a method together with its two variations for evaluating the customer's product-specific knowledge level. Section 4 shows our experiment results, and Section 5 concludes this paper after the discussions.

## 2. Research background

### 2.1. Product knowledge

In general, consumers can be assumed to have some amount of experience with or information about products they are using or plan to purchase. *Consumer knowledge* upon any particular product or service, i.e. product knowledge, has been traditionally referred to as *product familiarity* or *prior knowledge*. Alba and Hutchinson (1987) proposed that there were two major components, *familiarity* and *expertise*, in the construct of consumer knowledge, and *familiarity* and *expertise* were defined as 'the number of product-related experiences that have been accumulated by the consumer' and 'the ability to perform product-related tasks successfully,' respectively. They also indicated that increased product familiarity results in increased consumer expertise, different tasks require different types of expertise, and the successful performance of any particular task requires more than one type of knowledge.

There are many product knowledge-related differences existing between experts and non-experts, such as the most important and fundamental consumer behavior—decision making. Prior knowledge is not only related to consumers' behavioral motivation, such as information search intention, but facilitates the acquisition of new information and increases search efficiency (Brucks, 1985). The differences in information searching/processing combined with the differences in product knowledge produce different effects on consumer behavior. Alba and Hutchinson (1987) suggested that there should be significant differences between experts and non-experts in the size and composition of the set of alternatives they consider and in the nature of the attributes that are used to evaluate those alternatives. Rao and Monroe (1988) found that experts (i.e. those who have high product knowledge) and non-experts (i.e. those who have low product knowledge) are more likely to use product price for determining the quality of product, as compared to the group of consumers with intermediate product knowledge. Experts in a product domain may have different needs, regarding products and product-related information within that domain, from non-experts (Alba & Hutchinson, 1987). Since product knowledge may influence consumer's needs and behaviors, we propose that the knowledge of the customer's product knowledge level can be used to enhance the personalization strategy for EC companies.

Product knowledge can be further categorized into subjective knowledge, objective knowledge, and experience-based knowledge, where subjective knowledge measures a consumer's perception of how much she/he knows, objective knowledge measures what a consumer actually knows, and the experience-based knowledge measures the amount of purchasing or usage experience with the product (Brucks, 1985). It is pointed out that experience-based knowledge measures are less directly linked to behavior than are the other types of knowledge measures, subjective knowledge measures are conceptually and operationally distinct from objective knowledge measures, and subjective knowledge is closely related to one's behavioral motivation. Also according to Brucks (1985), not only is subjective knowledge easier to use but is closely related to confidence in one's decision making ability. In addition, Park and Lessig (1981) indicated that subjective knowledge serves as a better indicator of decision makers' systematic biases and heuristics than objective knowledge. Based on the above findings, we designed a system for automatically and non-intrusively assessing consumers' product knowledge. For the purpose of evaluating the performance of our proposed system, we also conducted questionnaire-based surveys to measure consumers' product knowledge, and the results from various approaches could be compared quantitatively.

### 2.2. Web usage mining

While data mining refers to extracting knowledge from a large amount of data (Han and Kamber, 2001), web usage mining can be viewed as a process of using data mining techniques to discover, extract, and analyze behavior patterns in web environment. Data mining is generally based on standard databases, but web usage mining is based on web-related data sources such as web logs, navigation paths, browsing patterns and others. Data mining techniques (such as cluster analysis, association rules mining, and sequential patterns mining) can be used for web usage mining (Changchien, et al., 2004; Chen, Park, & Yu, 1998; Chen, Han, & Yu, 1996). The most-widely used data sources in web usage mining are the web logs created by website servers to record how users access the site's web pages. Typically, a web-log records all HTTP request related information, which needs to go though a data preparation process before being applied to the pattern discovery process. The data preparation process recreates a user session file where each session is a collection of various requests from a user during a visit to the website (Cooley, Mobasher, & Srivastava, 1997). The pattern-discovery process derives knowledge about users in various forms including association rules, sequential patterns, usage clusters, and other implicit navigation behaviors (Mobasher, Cooley, & Srivastava, 2000), and such knowledge can be applied to various applications including web personalization, recommender systems, system improvement, business intelligence, and others (Srivastava, Cooley, Deshpande & Tan, 2000).

Web-usage mining based on web logs has some limitations caused by various issues related to proxy server, dynamic content, and local cache. Proxy servers are commonly used, in an enterprise or similar environment, to achieve the objective of performance improvement or security enforcement. All users accessing a website through the same proxy might be regarded as only one single user, since all requests would be logged as from the same IP address. In addition, some requests sent to a proxy might not be relayed to the website, if the same request has been made earlier and the web page of the response is readily available on the proxy. In this case, the proxy would never make a request to the website, therefore, there would be no log entries recorded on the website about this kind of client requests. Dynamic content is non-static information constructed by a program running on the server in response to user requests. A dynamic page with dynamic content can be created by using server side scripting technologies such as Microsoft's ASP (active server pages), Sun Microsystem's JSP (Java server pages), and Apache Software Foundation's PHP: Hypertext Preprocessor. Nowadays, most EC websites provide dynamic web pages because they are more powerful and flexible. However, it is possible to have multiple web pages (with different dynamic contents) accessed by users and logged by the website as multiple accesses to the same web page. In this case, web log based mining approach cannot distinguish which web page was actually rendered and delivered to users. Local cache is used by web browsers to keep web pages browsed in the past, for speeding up the response time of the requested web pages which are available in local cache and currently still up-to-date. Since this kind of requests would never reach web servers, it would not have any browsing record logged on the website. To resolve the above-mentioned issues, an approach of using information agents was proposed to record and track single-user surfing behaviors on the Internet continuously (Tu & Hsiang, 2000). However, this agent-based approach suffers the problem of sharing client machine by multiple users, since the agent cannot distinguish one from others for all users on the same computer. In addition, placing an agent program on client machines would face the scrutiny of security and privacy concerns.

To ameliorate the issues of the log based and the agent based web usage mining approaches, our research proposed a server based approach with a navigation recorder to track and record users' navigation information and browsing behaviors on a website. The navigation recorder is designed to work on an EC website offering dynamic content to make sure that every user request will be responded by the site with a freshly constructed dynamic page such that the local cache problem no longer exists. The server based design of our web mining approach automatically resolves the problems related to proxy server and dynamic content, since the navigation recorder on the website is capable of tracking all browsed web pages and contents in details. Furthermore, the non-intrusive nature of our server side tracking approach without planting agent programs onto client machines can dismiss, or at least alleviate, the litigations of security and privacy infringement.

### 2.3. Back propagation networks

Artificial neural network (ANN) models simulate the functions of nature's neurons networks by connecting the artificial neurons (Hagan, Demuth, & Beale, 1995).

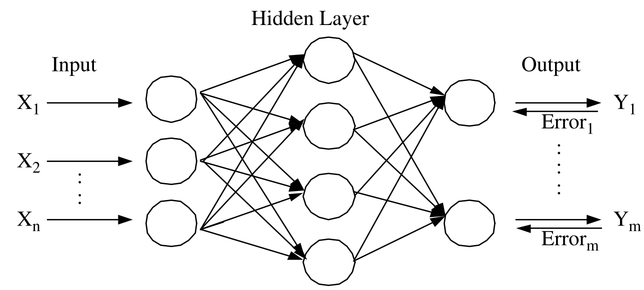


Fig. 1. Back Propagation Networks (BPN).

ANN users do not need to write complicated programs for solving their problems, since ANN is able to learn by itself and generate the result only based on input data. Back Propagation Network (BPN), one of the most well-known and widely used ANN models, is used in our research. As shown in Fig. 1, there are three layers in a BPN model: Input Layer, Hidden Layer, and Output Layer. Once the input information is entered into the Input Layer, Hidden Layer will start to compute and adjust the connection weights until the convergent results, as the outputs manifested in Output Layer, are obtained. The artificial neurons in Input Layer represent the variables without computing capability. The artificial neurons in Hidden Layer are hidden and responsible for processing the provided input data using transformation function. During training process, BPN keeps adjusting the connection weights based on the error between expected output and the actual value, and do not stop training until the network is convergent and becomes optimum. The basic mechanism in BPN is that the output values of all neurons will be multiplied by the corresponding weights to derive weighted values, which, will be summed up through the computing by using activation function, then generates the output signals. More details of BPN can be found in (Roth, 1990; Werbos, 1988).

BPN has been widely applied to various fields, including the analysis of signals of mineral deposits, the weather prediction on meteorology, diagnosis on the acute coronary embolism on medical science (Baxt, 1990), computer-aided design/manufacturing, stock price prediction (Baba & Kozaki, 1992), insurance examinations, futures trade (Bergerson & Wunsch, 1991), detecting credit card fraud (Rochester & Douglass, 1990), prediction of bankruptcy (Odom, 1989), tax analysis, etc. In our research, navigation patterns discovered by our server based web usage mining process are applied to BPN for automatically and non-intrusively detecting users' product knowledge levels.

## 3. Research methodology

### 3.1. The proposed system and its process flow

For enhancing the personalization strategy with an ultimate goal of attracting/retaining customers and enforcing the competitiveness for EC businesses, we built

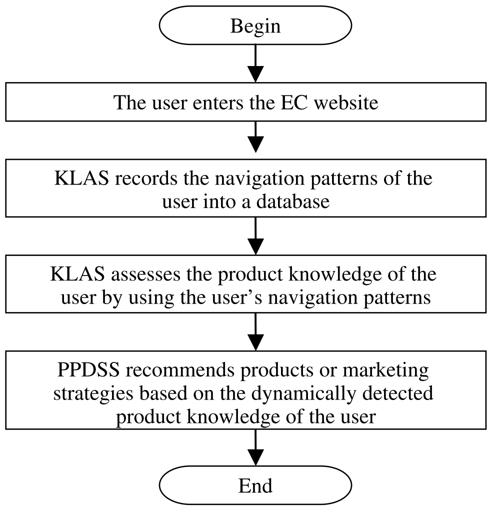


Fig. 2. The process flow of the integrated PPDSS and KLAS system.

a knowledge level assessment system (KLAS) to assess potential customer's product knowledge levels, and then integrated KLAS into an on-line personalized promotion decision support system (PPDSS). PPDSS was built, as one of our EC research projects, using data mining techniques in accordance with marketing strategies (for sales promotion and pricing) to help the business prepare the highly potential and suitable promotion products for each individual customer (Changchien, et al., 2004). Fig. 2 shows the process flow of the integrated PPDSS and KLAS system for achieving the goal of dynamic personalization.

When a user enters and starts a session on a website, KLAS would invoke its navigation recorder to start recording his/her browsing behavior and navigation path into a database. Immediately after the navigation path is recorded, KLAS would employ its built-in web mining techniques to match the path to pre-identified expert patterns and non-expert patterns. As shown in Fig. 3, these patterns are subdivided into different clusters based on the length of the navigation patterns, which are the sequences of web pages browsed by the user. In our initial design of KLAS, the pattern lengths ranging from 3 to 6 were adopted for mining navigation paths against both expert patterns and non-expert patterns. After the pattern mining process, KLAS will derive a collection of eight numbers (E3, E4, E5, E6, N3, N4, N5, N6) where E3, E4, E5, and E6 represent the discovered numbers of expert patterns whose lengths are 3, 4, 5, and 6, respectively, and similarly, N3, N4, N5, and N6 are the discovered numbers of non-expert patterns whose lengths are 3, 4, 5, and 6, respectively. This collection of eight numbers is used by KLAS as the input to its built-in BPN model for detecting the user's product knowledge level, which in turn would be used by PPDSS for dynamically recommending products and presenting product information to the user according to the marketing (sales promotion and pricing) strategies of the EC business.

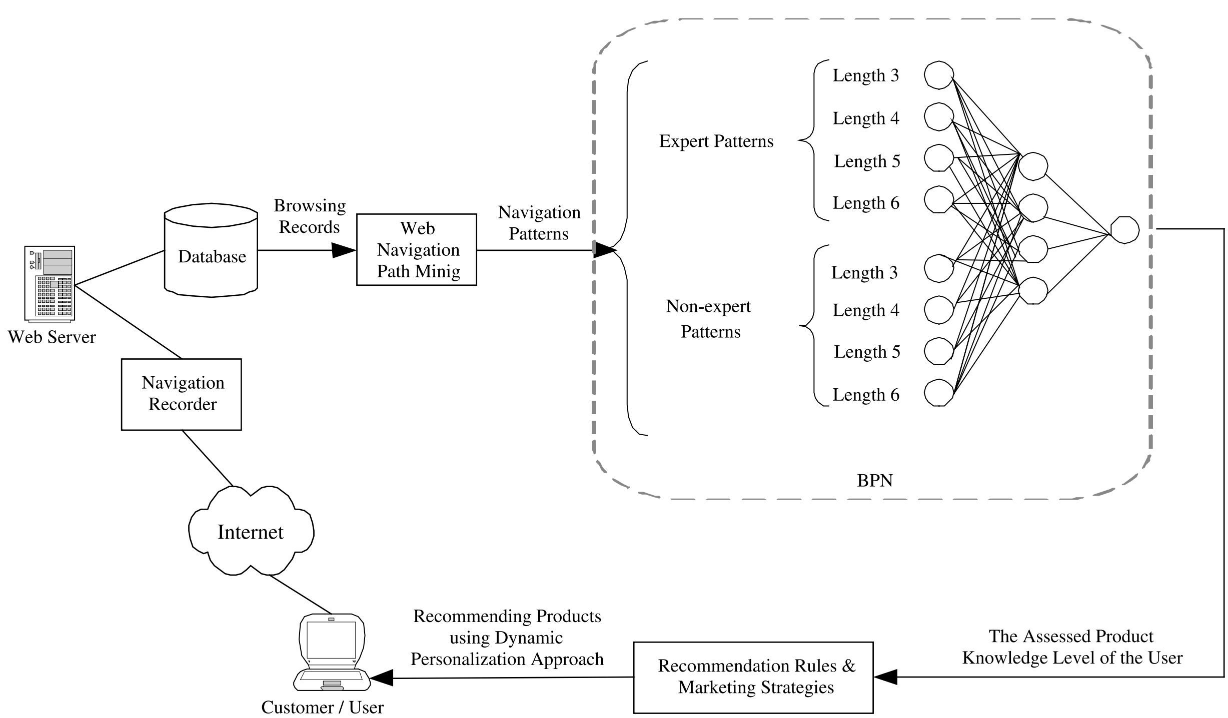


Fig. 3. The architecture of the navigation pattern based KLAS system.

### 3.2. BPN model training

While the integrated PPDSS and KLAS system can assess the user's product knowledge level and achieve the goal of dynamic and personalized recommendation and marketing in a real-time on-line web environment, a pre-defined and well trained BPN model is needed in this proposed system. BPN model training process is highly important, since it affects the accuracy of the knowledge assessment, i.e. the quality and applicability of KLAS. This model training process is required for every new product category and for every new website structure, and it is divided into four steps: collecting training data, mining web navigation paths, extracting expert and non-expert patterns, and training the BPN networks. The first step is to collect and record navigation patterns, by identifying and inviting both experts and non-experts to participate in our research and ask them to browse through the designated website. Afterwards, the users' navigation paths will be analyzed using applicable web usage mining algorithms to extract those more significant and meaningful navigation patterns, which would be further categorized into expert patterns and non-expert patterns. An important task of this categorization step is to remove the common patters, which manifest and present in both expert group and non-expert group, such that the remaining patterns in both groups could more characteristically distinguish experts from non-experts. Finally, expert patterns and non-expert patterns derived from previous three steps will be used to create training datasets, which contain collections of eight numbers as described in the previous sub-section. These numbers, representing the numbers of discovered expert and non-expert patterns with various lengths ranging from 3 to 6, would be used as the input to train the BPN model.

### 3.3. The sample application

The application to the proposed product knowledge level assessment system is illustrated with an on-line store. To prove the concept, an experiment EC website was built in

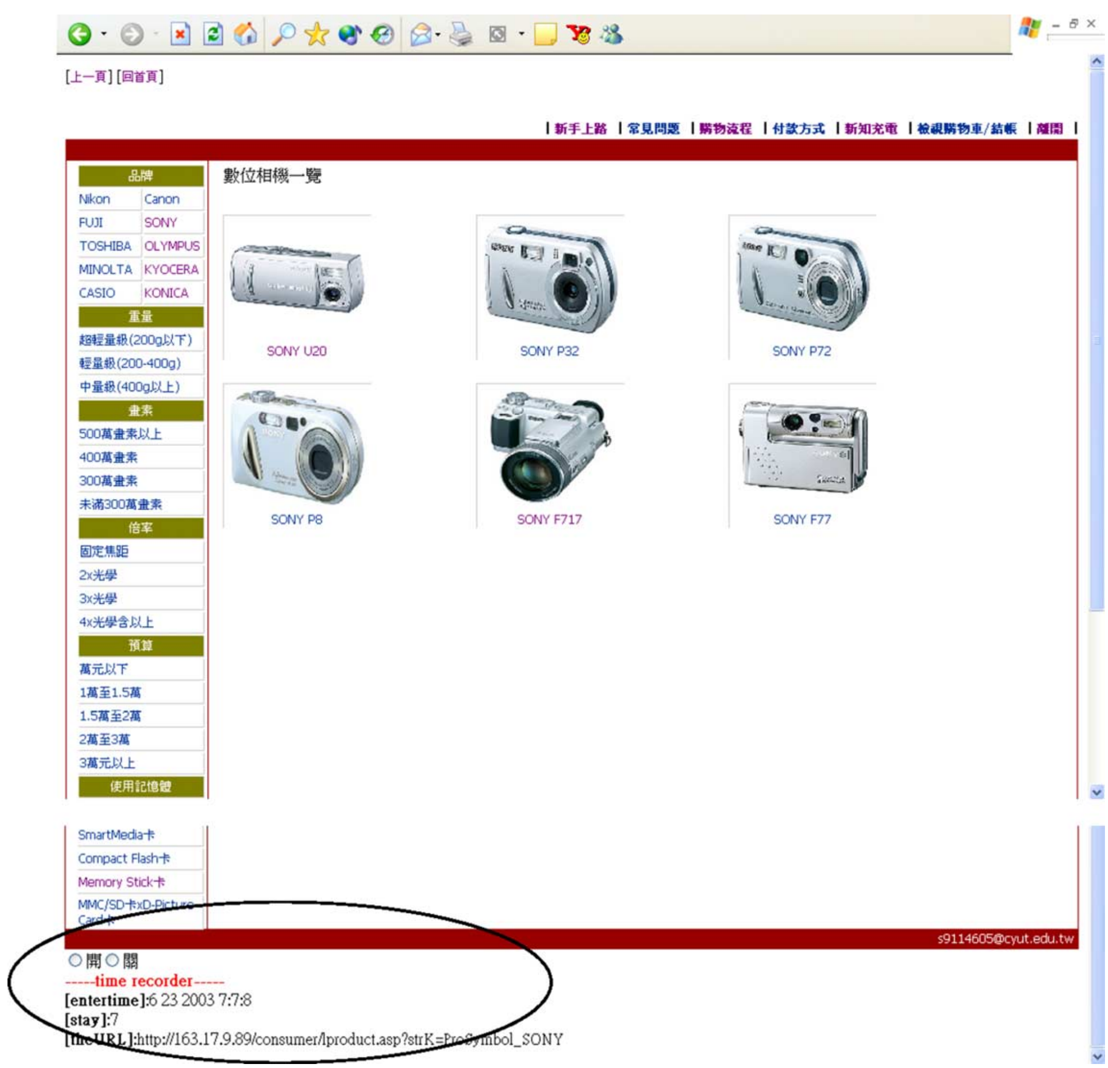


Fig. 4. A sample web page of our experiment website selling digital cameras.

our research to simulate such an on-line store in conducting B2C business of selling digital cameras. Other than the necessary components of a typical B2C website, a server-based navigation recorder for tracking and recording users' navigation behavior, as described in previous section, is a core component for facilitating the task of navigation path mining. A sample web page of our experiment EC website is shown in Fig. 4, in which the information recorded by the navigation recorder is displayed at the bottom.

For training the BPN model of KLAS, there are 46 experts and 119 non-experts identified through our selection process to browse through our digital camera website. For selecting experts, we went to digital camera related news groups and chat-rooms on the Web, and invited those users who frequently provided opinions or useful information to other users as experts as long as they also regarded themselves as digital camera experts. In this selection approach, we actually rely on the concept of subjective knowledge, i.e. all selected experts consider themselves to possess enough knowledge about digital cameras. For the selection of non-experts, we also use the similar subjective knowledge approach, i.e. the selected non-experts are those who do not regard themselves as digital camera experts.

When a user is browsing through various web pages on the site, the navigation recorder would track and record the browsing sequence of the user into a database table, and an example of this table is presented in Table 1, which actually represents only a small portion of the real navigation information recorded. In this table either userID or userIP can be used to identify each user, while sessionID can be used to distinguish every browsing sequence. For example, it is obvious to see that Table 1 contains three browsing sequences of two users, one sequence for taco and two sequences for wice. The theURL field records the actual web page a particular user visits. For the web page identified by theURL, the TimeIn field and the TimeOut field track the entrance time and the exit time respectively. The last column (named D) in Table 1 represents the duration of the stay on each particular web page (i.e. D= TimeOut—TimeIn). It is believed that the value of D may serve as an indicator of how dedicated the user is 'working' on the web page, and it would be interesting to find out whether and how the value of D would affect the result of our proposed product knowledge level assessment system.

The data in Table 1, or more specifically the recorded navigation paths of users, need to go through a pre-processing task similar to those methods presented by Cooley et al. (1997) for building a server session file. To facilitate this task, the recorded navigation paths are encoded by using the URL-to-code mapping scheme described in Table 2. Although row 2 and row 3 in Table 1 share the same URL (http://163.17.9.89/consumer/lproduct.asp), they actually refer to different dynamic pages and their dynamic parameters (strK=ProSymbol_Nikon and strK=ProSymbol_Canon) recorded together with the URL are different. Therefore, dynamic pages under the same URL but with different parameter values are treated as different web pages. An effective and efficient data structure for representing navigation paths/sequence can be derived using this encoding scheme. For example, according to the coding scheme defined by Table 2, a navigation path/sequence containing 9 pages may be encoded as a path/sequence of '**044**000**012**037**035**034**032**031**004**,' where each page is represented by a code of three digits. Through a navigation pattern mining process, this navigation path can be used to derive navigation patterns with various lengths ranging from 3 to 6. The navigation patterns extracted from this example are shown in Table 3. After this

Table 1
An example of the recorded user navigation information

| No | userID | sessionID | theURL | userIP | TimeIn | TimeOut | D |
|---|---|---|---|---|---|---|---|
| 1 | Taco | 1050883611 | http://163.17.9.8/lproduct.asp | 211.74.236.25 | 03/6/3pm08:51:30 | 03/6/3pm08:51:56 | 26 |
| 2 | Taco | 1050883611 | http://163.17.9.8/lproduct.asp?strK=K_2 | 211.74.236.25 | 03/6/3pm08:51:57 | 03/6/3pm08:52:23 | 26 |
| 3 | Taco | 1050883611 | http://163.17.9.8/lproduct.asp?strK=P_K | 211.74.236.25 | 03/6/3pm08:52:24 | 03/6/3pm08:52:33 | 9 |
| 4 | Taco | 1050883611 | http://163.17.9.8/logout.asp | 211.74.236.25 | 03/6/3pm08:52:33 | 03/6/3pm08:52:38 | 5 |
| 5 | Wice | 1050883677 | http://163.17.9.8/lproduct.asp | 163.17.9.75 | 03/6/4pm05:33:51 | 03/6/4pm05:34:11 | 20 |
| 6 | Wice | 1050883677 | http://163.17.9.8/lproduct.asp?strK=P_N | 163.17.9.75 | 03/6/4pm05:34:11 | 03/6/4pm05:34:12 | 1 |
| 7 | Wice | 1050883677 | http://163.17.9.8/lproduct.asp?strK=P_C | 163.17.9.75 | 03/6/4pm05:34:13 | 03/6/4pm05:34:17 | 4 |
| 8 | Wice | 1050883677 | http://163.17.9.8/a_new.asp | 163.17.9.75 | 03/6/4pm05:34:18 | 03/6/4pm05:34:25 | 7 |
| 9 | Wice | 1050883682 | http://163.17.9.8/lproduct.asp?strK=P_C | 163.17.9.75 | 03/6/5pm03:25:20 | 03/6/5pm03:25:33 | 13 |
| 10 | Wice | 1050883682 | http://163.17.9.8/lproduct.asp?strK=K_1 | 163.17.9.75 | 03/6/5pm03:25:34 | 03/6/5pm03:25:38 | 4 |
| 11 | Wice | 1050883682 | http://163.17.9.8/lproDtl.asp?ProNo=45 | 163.17.9.75 | 03/6/5pm03:25:49 | 03/6/5pm03:25:52 | 3 |
| 12 | Wice | 1050883682 | http://163.17.9.8/lproduct.asp?strK=P_C | 163.17.9.75 | 03/6/5pm03:25:52 | 03/6/5pm03:25:56 | 4 |
| 13 | Wice | 1050883682 | http://163.17.9.8/lproDtl.asp?ProNo=44 | 163.17.9.75 | 03/6/5pm03:25:56 | 03/6/5pm03:36:00 | 4 |
| 14 | Wice | 1050883682 | http://163.17.9.8/lproDtlP.asp?ProNo=44 | 163.17.9.75 | 03/6/5pm03:26:00 | 03/6/5pm03:36:02 | 2 |
| 15 | Wice | 1050883682 | http://163.17.9.8/lproDtlW.asp?ProNo=44 | 163.17.9.75 | 03/6/5pm03:26:03 | 03/6/5pm03:36:04 | 1 |
| 16 | Wice | 1050883682 | http://163.17.9.8/lproDtlF.asp?ProNo=44 | 163.17.9.75 | 03/6/5pm03:26:04 | 03/6/5pm03:36:06 | 2 |
| 17 | Wice | 1050883682 | http://163.17.9.8/lproDtlG.asp?ProNo=44 | 163.17.9.75 | 03/6/5pm03:26:07 | 03/6/5pm03:36:10 | 3 |
| 18 | Wice | 1050883682 | http://163.17.9.8/lbasket.asp?ProNo=44 | 163.17.9.75 | 03/6/4pm05:36:10 | 03/6/4pm05:36:12 | 2 |
| 19 | Wice | 1050883682 | http://163.17.9.8/lpay.asp | 163.17.9.75 | 03/6/4pm05:36:12 | 03/6/4pm05:36:18 | 6 |
| 20 | Wice | 1050883682 | http://163.17.9.8/lpayOk.asp | 163.17.9.75 | 03/6/4pm05:36:18 | 03/6/4pm05:36:21 | 3 |

Table 2
The encoding table used to encode the navigation paths

| No | file_detail | TheUrl | code |
|---|---|---|---|
| 1 | Overview: products | http://163.17.9.89/consumer/lproduct.asp | 000 |
| 2 | By brand—Nikon | http://163.17.9.89/consumer/lproduct.asp?strK=ProSymbol_Nikon | 001 |
| 3 | By brand—Canon | http://163.17.9.89/consumer/lproduct.asp?strK=ProSymbol_Canon | 002 |
| 4 | By brand—FUJI | http://163.17.9.89/consumer/lproduct.asp?strK=ProSymbol_FUJI | 003 |
| 5 | By brand—SONY | http://163.17.9.89/consumer/lproduct.asp?strK=ProSymbol_SONY | 004 |
| 6 | By brand—TOSHIBA | http://163.17.9.89/consumer/lproduct.asp?strK=ProSymbol_TOSHIBA | 005 |
| 7 | By brand—OLYMPUS | http://163.17.9.89/consumer/lproduct.asp?strK=ProSymbol_OLYMPUS | 006 |
| 8 | By brand—MINOLTA | http://163.17.9.89/consumer/lproduct.asp?strK=ProSymbol_MINOLTA | 007 |
| 9 | By brand—KYOCERA | http://163.17.9.89/consumer/lproduct.asp?strK=ProSymbol_KYOCERA | 008 |
| 10 | By brand—CASIO | http://163.17.9.89/consumer/lproduct.asp?strK=ProSymbol_CASIO | 009 |
| 11 | By brand—KONICA | http://163.17.9.89/consumer/lproduct.asp?strK=ProSymbol_KONICA | 010 |
| 12 | By weight—light | http://163.17.9.89/consumer/lproduct.asp?strK=Kheavy_1 | 011 |
| 13 | By weight—common | http://163.17.9.89/consumer/lproduct.asp?strK=Kheavy_2 | 012 |
| | ⋮ | ⋮ | ⋮ |
| 472 | Move product from market basket | http://163.17.9.89/consumer/dbasket.asp?ProNo=13 | 471 |
| 473 | Pay up | http://163.17.9.89/consumer/lpay.asp | 472 |
| 474 | Finish pay up | http://163.17.9.89/consumer/lpayOk.asp | 473 |

pattern extracting process is applied to the navigation paths recorded for both experts and non-experts, the common patterns shared by both the expert group and the non-expert group need to be discarded, such that the remaining expert patterns and non-expert patterns would become more powerful in distinguishing experts from non-experts.

As illustrated in Fig. 3, after the extracted navigation patterns are categorized into various lengths from 3 to 6, KLAS will derive a collection of eight numbers, representing the discovered numbers of expert patterns and non-expert patterns with various lengths from 3 to 6, as the input dataset to train the built-in BPN model of KLAS. Once the training process is completed, i.e. convergent results/outputs manifested in Output Layer are obtained, KLAS would be ready for detecting the user's product knowledge level.

### 3.4. Performance evaluations

While the subjective knowledge measures a consumer's perception of how much she/he knows, objective knowledge measures what a consumer actually knows. The accuracy of the proposed KLAS (shown in Fig. 3) on detecting customer's product knowledge levels is compared against measures based on the concept of subjective knowledge, since all expert navigation patterns in our experiment dataset were extracted from the navigation paths of those participants who regarded themselves as digital camera experts.

For the purpose of comparing the performances between the conventional approach of using questionnaire and the proposed innovative approach of navigation pattern mining, a questionnaire was used to create a benchmark measure of the customer's product knowledge level. The questionnaire was designed to ask questions in three categories: usage experience, basic concept, and advanced knowledge. The survey results/scores derived from the questionnaire can be treated as the indicators of the customer's objective knowledge measure, which is conceptually and operationally distinct from the subjective knowledge measure (Brucks, 1985). Instead of using the scores of questionnaire directly, an approach of using both BPN model and questionnaire was used in our research to detect the customer's product knowledge level. The accuracy of this approach is compared against the measures based on the concept of subjective knowledge as well. The architecture of the questionnaire based knowledge level assessment system is shown in Fig. 5.

Finally, as shown in Fig. 6, an enhancement to KLAS using hybrid inputs from both extracted navigation patterns and questionnaire answers was created to conduct an interesting experiment for finding out whether this hybrid approach would result in a better performance in terms of the accuracy on detecting the customer's product knowledge level. Again, the accuracy of this hybrid approach is

Table 3
An example of extracting navigation patterns from a navigation path

| Navigation | **044**000**012**037**035**034**032**031**004** (containing 9 web pages) |
|---|---|
| Extracted length 3 patterns | 044000012 000012037 012037035 037035034 035034032 034032031 032031004 |
| Extracted length 4 patterns | 044000012037 000012037035 012037035034 037035034032 035034032031 034032031004 |
| Extracted length 5 patterns | 044000012037035 000012037035034 012037035034032 037035034032031 035034032031004 |
| Extracted length 6 patterns | 044000012037035034 000012037035034032 012037035034032031 037035034032031004 |

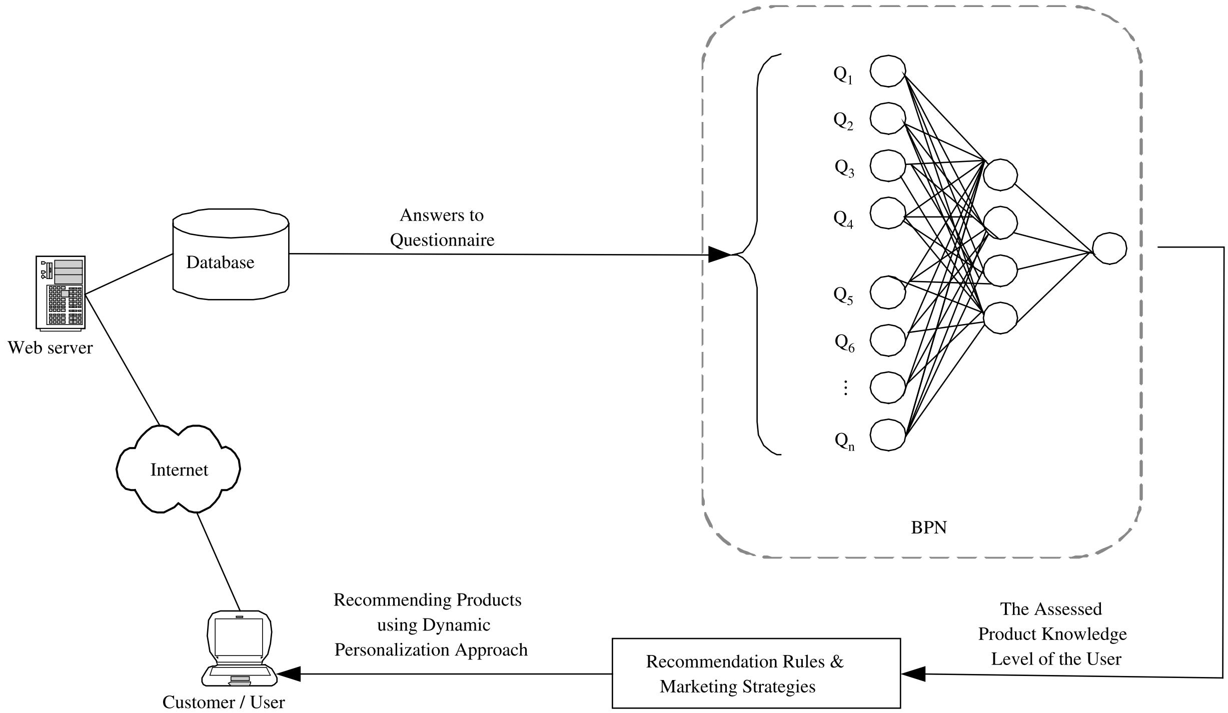


Fig. 5. The architecture of the questionnaire based KLAS system.

compared against the measures of the customer's subjective knowledge level.

## 4. Experiment results and performance evaluation

### 4.1. Experiment I—using navigation patterns

The shopping website is equipped with a navigation recorder for tracking each customer's navigation path and extracting his/her navigation patterns, and the navigation patterns extracted from all 165 participants are derived and stored by KLAS as the experiment dataset. The patterns in the dataset, as shown in Fig. 3, will be used to construct inputs to the BPN model of KLAS for either training the BPN model or assessing the customer's knowledge level on digital cameras. 42 experts and 108 non-experts from the dataset were randomly selected for training the BPN, and the rest of input data of 4 experts and 11 non-experts would be used as test data for evaluating the performance of the trained BPN model. The normalized output value, $v$, derived from the previously trained BPN model could be used to determine whether the customer is an expert (when $v$ is higher than 0.5) or non-expert (when $v$ is less than 0.5). This

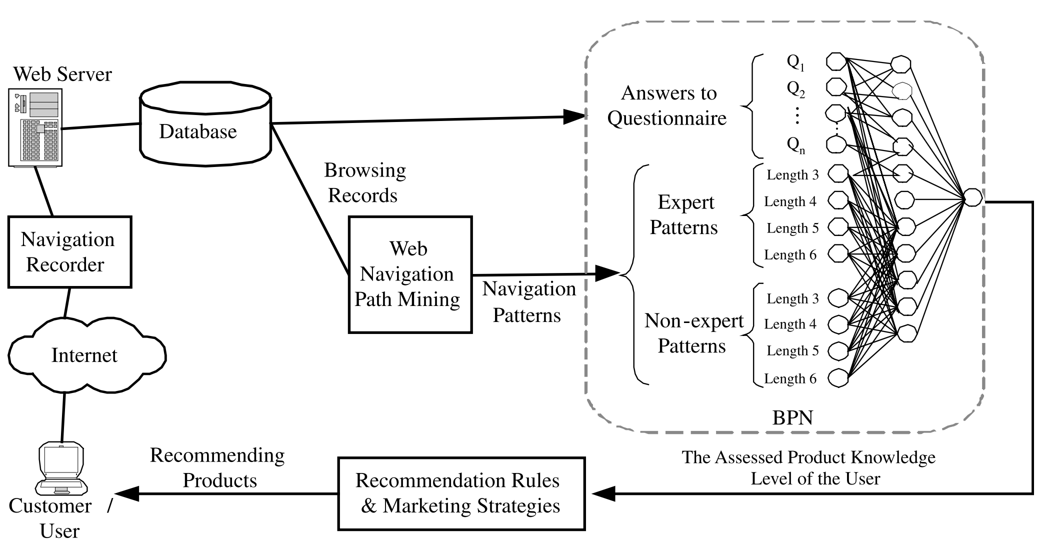


Fig. 6. The architecture of the hybrid KLAS system.

experiment was performed 10 times with various randomly selected and mutually exclusive training data sequences and test data sequences, and an average accuracy rate of 78%, as compared against the concept of subjective knowledge level, was derived from experiment I.

### 4.2. Experiment II—using questionnaire

The questionnaire-based approach was adapted and incorporated into the proposed KLAS system for conducting experiment II as shown in Fig. 5. All 165 participants were asked to complete the on-line questionnaire, and their answers were collected and stored in KLAS as the second experiment dataset, which was used to construct inputs to the BPN model of the questionnaire based KLAS for either training the BPN model or assessing the customer's knowledge level on digital cameras. Using the same experiment procedures, 42 experts and 108 non-experts from the dataset were randomly selected for training the BPN, and the rest of input data of 4 experts and 11 non-experts would be used as test data for evaluating the performance of the trained BPN model. An average accuracy rate of 84%, as compared against the concept of subjective knowledge level, was derived from experiment II.

### 4.3. Experiment III—using combined navigation patterns and questionnaire

As shown in Fig. 6, the navigation patterns as well as the answers to the questionnaire were used as the input to the BPN model of the hybrid KLAS system. The performance of this hybrid approach derived from experiment III, under the same procedures as experiments I and II, achieved an average accuracy rate of 93%, which was better than the ones of the pure navigation pattern based approach (78%) and the pure questionnaire based approach (84%).

## 5. Conclusion and discussion

It is quite a challenge that a business faces more competitors in Internet than in traditional market, and the customer's loyalty in the Internet is low compared with traditional market so that it is a difficult problem for a business to attract and retain customers in EC. Traditional mass marketing is no longer effective for EC in the Internet, and thus more precise on-line one-to-one marketing for better suiting each customer becomes more and more important for competing on the Internet, along with the use of highly advanced data analysis techniques and the development of new marketing strategies for EC. Hence, an on-line personalized promotion decision support system (PPDSS) was developed in our previous research (Changchien,et al., 2004), to assist a business in intelligently developing the marketing strategy and the on-line one-to-one promotion products based on the experiences analyzed and retrieved from the historical transactions. In this paper, an innovative non-intrusive method of assessing the customer's product knowledge level, which can be used to enhance the effectiveness of personalization strategy in an EC environment, was proposed and implemented through our knowledge level assessment system (KLAS). Since KLAS was designed to ameliorate the constrains of the log based and the agent based web usage mining techniques by using a server-side approach to track and record users' browsing behaviors on the website, it could be more effectively and proactively incorporated into various applications in gaining competing advantage for EC companies.

For the purpose of evaluating the performance of our proposed system, we also enhanced KLAS to accept questionnaire-based input so that not only could the results from various approaches be compared quantitatively, but the navigation pattern based KLAS (see Fig. 3) and the questionnaire-based KLAS (see Fig. 5) could be combined to derive a hybrid KLAS (see Fig. 6). It was shown from our experiments that the proposed navigation pattern based KLAS achieved an accuracy rate of 78% in detecting whether a user is an expert or not, the benchmark derived from the questionnaire-based KLAS offered a better accuracy rate of 84%, and the hybrid KLAS delivered the best accuracy rate of 93%. It is good to see that the proposed navigation pattern based KLAS is a promising way of dynamically, proactively, and non-intrusively detecting users' product-specific knowledge levels, given that its accuracy rate derived from our experiments is good and only slightly lower than the accuracy rate of the questionnaire-based KLAS, which, however, is intrusive and may not be accepted by at least some users. However, for those users willing to complete the questionnaire, the proposed navigation pattern approach can be integrated with the questionnaire-based approach to create a hybrid KLAS which can achieve a significantly improved accuracy rate (93% based on our experiments).

In addition to learning and constructing customer profiles and preferences (i.e. to know about the users), EC businesses may use KLAS to further enhance their personalization strategy by understanding the customer's knowledge level about each target product (i.e. to know about what the users know about the product). This domain knowledge of the customer's product-specific knowledge level can be applied to various areas (such as customer segmentation, customer relationship management, call center, web personalization, recommendation system, etc.) in conducting EC business. For example, in a one-to-one marketing website, the information content, information organization, and information presentation about a promotion item can be tailored according to various knowledge levels of customers on the target item. Knowing a user's product-specific knowledge level has the potential to confer considerable competing advantage in the development of 'dynamic personalization' strategy for EC companies. The capability and methodology of assessing user's product-specific knowledge quickly and effectively can facilitate the design of an EC website for dynamic personalization, which in turn can be used to facilitate the customer orientation design. Customer orientation always emphasizes on customer's interest, and it stresses the derivation of customer profiles and the construction of customer knowledge base for enterprises as an intangible asset that is difficulty to be imitated by competitors (Deshpande, Farley, & Webster, 1993). Through the customer orientation strategy to leverage customer knowledge, enterprises can avoid competing on pricing and flexibly provide differential prices based on the customer's demand curve, and eventually get higher average prices (Roberts, 2000). Our proposed KLAS based dynamic personalization is a feasible and attractive tool for today's enterprises to enhance the effectiveness of customer orientation and create strategic advantage by raising the entrance barriers. To prove the concept, not only was KLAS integrated into our PPDSS to show how KLAS can achieve the goal of dynamic personalization and customer orientation design (see Fig. 2), but an empirical study was conducted using a KLAS application website for selling digital cameras.

Although we have proposed an innovative, non-intrusive, and promising method for assessing users' product-specific knowledge, there are some constraints on its general applicability. First, each implementation of this method would only work on one category of EC products. For an EC website selling various categories of products, the complexity and the degree of difficulty in design and implementation of the KLAS system would significantly increase. Second, KLAS relies on its BPN model for detecting whether a user is an expert, and the BPN model needs to be trained before KLAS can be put in use. However, the BPN training process requires pre-defined and mutually exclusive expert patterns and non-expert patterns, as well as a training dataset collected in advance. The accuracy of KLAS is highly dependent on the quality and the quantity of the training dataset. Third, this method may only work for products with some specific characteristics and attributes. Since digital camera was selected as the sample product in our study, it is highly possible that KLAS can be applied to other 3C products, but it is still uncertain whether KLAS can be used by websites selling non-3C products. Some pretests or further empirical studies may need to be conducted to classify the candidate products before incorporating KLAS into an EC website, and it is clear that a consistent measurement for product categorization needs to be established for using the proposed KLAS.

## References

Aggarwal, C. C., Wolf, J. L., Wu, K. L., & Yu, P. S. (1999). *Horting hatches an egg: A new graph-theoretic approach to collaborative filtering Proceedings of ACM KDD-99, international conference on knowledge discovery and data mining* pp. 201–212.

Alba, J. W., & Hutchinson, J. W. (1987). Dimensions of consumer expertise. *Journal of Consumer Research*, *13*(4), 411–454.

Allen, C., Kania, D., & Yaeckel, B. (1998). *Internet world guide to one-to-one Web marketing*. New York: Wiley.

Baba, N., & Kozaki, M. (1992). An intelligent forecasting system of stock price using neural networks. *IJCNN-92*, *I*, 371–377.

Baxt, W. G. (1990). Use an artificial neural network for data analysis in clinical decision-making: The diagnosis for acute coronary occlusion. *Neural Computation*, *2*, 480–489.

Bergerson, K., & Wunsch, D. C. (1991). A commodity trading model based on a neural network-expert system hybrid. *IJCNN-91*, *I*, 289–293.

Brucks, M. (1985). The effect of product class knowledge on information search behavior. *Journal of Consumer Research*, *12*(1), 1–16.

Changchien, S. W., Lee, C. F., & Hsu, Y. J. (2004). On-line personalized sales promotion in electronic commerce. *Expert Systems with Applications*, *27*(1), 35–52.

Chen, M.-S., Han, J., & Yu, P. S. (1996). Data mining: An overview from a database perspective. *IEEE Transactions on Knowledge and Data Engineering*, *8*(6), 866–883.

Chen, M.-S., Park, J. S., & Yu, P. S. (1998). Efficient data mining for path traversal patterns. *IEEE Transactions on Knowledge and Data Engineering*, *10*(2), 209–222.

Cooley, R., Mobasher, B., & Srivastava, J. (1997). *Grouping web page references into transactions for mining world wide web browsing patterns Knowledge and data engineering workshop, Newport Beach, CA* pp. 2–9.

Deshpande, R., Farley, J. U., & Webster, F. E. (1993). Corporate culture, customer orientation, and innovativeness in Japanese firms: A quadrad analysis. *Journal of Marketing*, *57*(1), 23–37.

Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, *35*(12), 61–70.

Hagan, M. T., Demuth, H. B., & Beale, M. (1995). *Neural network design PWS publishing company*.

Han, J., & Kamber, M. (2001). *Data mining: Concepts and techniques*. Los Altos, CA: Morgan Kaufmann.

Kim, J. K., Cho, Y. H., Kim, W. J., Kim, J. R., & Suh, J. H. (2002). A personalized recommendation procedure for Internet shopping support. *Electronic Commerce Research and Applications*, *1*(3–4), 301–313.

Konstan, J. A., Miller, B. N., Maltz, D., Herlocker, J. L., Gordon, L. R., & Riedl, J. (1997). GroupLens: Applying collaborative filtering to Usenet news. *Communications of the ACM*, *40*(3), 77–87.

Krulwich, B., & Burkey, C (1996). Learning user information interests through extraction of semantically significant phrases. Proceedings of the AAAI spring symposium on machine learning in information access..

Lang, K. (1995). *Newsweeder: Learning to filter netnews Proceedings of the 12th international conference on machine learning* pp. 331–339.

Lee, C.-H., Kim, Y.-H., & Rhee, P.-K. (2001). Web personalization expert with combining collaborative filtering and association rule mining technique. *Expert Systems with Applications*, *21*(3), 131–137.

Lee, W. P., Liu, C. H., & Lu, C. C. (2002). Intelligent agent-based systems for personalized recommendations in internet commerce. *Expert Systems with Applications*, *22*(4), 275–284.

Li, Y., Lu, L., & Xuefeng, L. (2005). A hybrid collaborative filtering method for multiple-interests and multiple-content recommendation in E-commerce. *Expert Systems with Applications*, *28*(1), 67–77.

Li, Y., & Zhong, N. (2004). Web mining model and its applications for information gathering. *Knowledge-Based Systems*, *17*(3), 207–217.

Lin, J., & Lu, H. (2000). Towards an understanding of the behavioural intention to use a web site. *International Journal of Information Management*, *20*(3), 197–208.

Liu, C., & Arnett, K. (2000). Exploring the factors associated with web site success in the context of electronic commerce. *Information and Management*, *38*(1), 23–33.

Mobasher, B., Cooley, R., & Srivastava, J. (2000). Automatic personalization based on Web usage mining. *Communications of the ACM*, *43*(8), 142–151.

Odlyzko, A. M. (2003). Internet traffic growth: Sources and implications. In B. B. Dingel, W. Weiershausen, A. K. Dutta, & K.-I. Sato, *Optical transmission systems and equipment for WDM networking II*. *Proceedings of SPIE*, 1–15.

Odom, M. D. (1989). A neural network model for bankruptcy prediction. *IJCNN-89*, *II*, 163–168.

Park, C. W., & Lessig, V. P. (1981). Familiarity and its impact on consumer decision biases and heuristics. *Journal of Consumer Research*, *8*(9), 223–230.

Rao, A. R., & Monroe, K. B. (1988). The moderating effect of prior knowledge on cue utilization in product evaluations. *Journal of Consumer Research*, *15*(2), 253–264.

Roberts, J. H. (2000). Developing new rules for new markets. *Journal of the Academy of Marketing Science*, *28*(1), 31–44.

Rochester, J. B., & Douglass, D. P. (1990). New business for neurocomputing. *I/S Analyzer*, *28*(2), 1–5.

Roth, M. (1990). Survey of neural-network technology for automatic target recognition. *IEEE Transaction on Neural Networks*, *1*, 28–43.

Rucker, J., & Polenco, M. J. (1997). Siteseer: Personalized navigation for the web. *Communications of the ACM*, *40*(3), 73–76.

Shardanand, U., & Maes, P. (1995). *Social information filtering: Algorithm for automating word of mouth Proceedings of international conference on human factors in computing systems* pp. 210–217.

Srivastava, J., Cooley, R., Deshpande, M., & Tan, P. (2000). Web usage mining: Discovery and applications of usage patterns from Web data. *SIGKDD Explorations*, *1*(2), 1–12.

Terveen, L., Hill, W., Armento, B., McDonald, D., & Creter, J. (1997). PHOAKS: A system for sharing recommendations. *Communications of the ACM*, *40*(3), 59–62.

Tu, H. C., & Hsiang, J. (2000). An architecture and category knowledge for intelligent information retrieval agents. *Decision Support Systems*, *28*(3), 255–268.

Wang, F.H., & Thao, S.M. (2003). A study on personalized Web browsing recommendation based on data mining and collaborative filtering technology. Proceedings of national computer symposium, Taiwan, pp. 18-25.

Wang, F. H., & Shao, H. M. (2004). Effective personalized recommendation based on time-framed navigation clustering and association mining. *Expert Systems with Applications*, *27*(3), 365–377.

Weng, S. S., & Liu, M. J. (2004). Feature-based recommendations for one-to-one marketing. *Expert Systems with Applications*, *26*(4), 493–508.

Werbos, P. J. (1988). Generalization of backpropagation with application to gas market model. *Neural Networks*, *1*, 339–356.

Wind, J., & Rangaswamy, A. (2001). Customerization: The next revolution in mass communication. *Journal of Interactive Marketing*, *15*(1), 13–32.

Yang, C. C., Yen, J., & Chen, H. C. (2000). Intelligent internet searching agent based on hybrid simulated annealing. *Decision Support Systems*, *28*(3), 269–277.

Yu, P. S. (1999). *Data mining and personalization technologies Proceedings of the sixth international conference on database systems for advanced applications* pp. 6–13.