早稲田大学審査学位論文
博士（人間科学）

User-Centric Recommendation Based on Gradual Adaptation Model and Behavior Analysis

挙動解析と逐次適応モデルに基づくユーザ中心の推薦手法

２０１２年７月

早稲田大学大学院　人間科学研究科

陳　健

CHEN, Jian

研究指導教員：　金　群　教授

早稲田大学大学院人間科学研究科博士学位審査論文

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

Abstract

Information seeking is a kind of behavior that people attempt to obtain information, which related to, but yet different from information retrieval. Generally, information retrieval focuses on retrieving information from resource effectively and efficiently, but information seeking pays attention to the process of how to retrieve information from resource, in which the successful process of information seeking enable people to apperceive the intangible knowledge. In other words, people can perceive in terms of a past experience. Further, the successful experiences are beneficial to effectively and efficiently improve the information retrieval better.

Currently, more and more people seek wanted information from Web resource. In order to provide an effective and efficient environment to help people obtain their wanted information from Web resource, many search engine providers chose delivering a service namely personalized search (e.g. Google Personalized Search[1]). The Google Personalized Search service takes into account the website access histories of users, which neglects the similarities of users. Therefore, the effectiveness of improving search results is limited. In the another way, information recommendation is regarded as an approach based on information filtering, which seeks to predict the ‘rating’ or ‘preference’ of an item to users based on the access history or social element[2].

Differing from many studies of recommendation that provided the final results directly, our study focuses on recommending an optimized process of information seeking to people by analyzing information seeking behavior. Based on process mining, we propose an integrated adaptive framework to support and facilitate

---

[1] http://en.wikipedia.org/wiki/Google_Personalized_Search
[2] http://en.wikipedia.org/wiki/Recommender_system

individual recommendation based on *Gradual Adaptation* model (GA model) [1] that gradually adapts to a target user's transition of needs and behaviors of information access, including various search related activities, over different time spans. Successful processes of information seeking are extracted from the information seeking histories of users and used to create a dynamic Bayesian network of seeking actions. According to the posterior probability of seeking actions, the successful processes of information seeking are optimized as a series of action units for the target users whose information access behavior patterns are similar to the reference users. And then, the optimized processes of information seeking are navigated to the target users according to their transition of interest focus.

In this thesis, in addition to describing some definitions and measures introduced in this study, we present a GA model and show how the model is applied to detect the transition of users' needs, and then, we try to extract behavior patterns from learning process that is regarded as a specific seeking process. Our analysis results show that most of users often use their preferred learning patterns in their learning processes, and the learning achievement is affected by the learning process. Based on these findings, we attempt to optimize the learning process using the extracted learning patterns, infer the learning goal of target users, and provide a goal-driven navigation of individualized learning process according to the similarity of the extracted learning patterns. On this basis, we propose an *Optimized Process Recommendation* model (OPR model) and show the system architecture. Finally, we conclude the thesis with the summary of the proposed approach and outlines future works.

1 Introduction

1.1 Background

1.2 Purpose of this Study

1.3 Organization of the Thesis

In this chapter, we introduce the background of information recommendation and information behavior access behavior analysis at first. The motivation and purpose of this study are discussed at second. After them, the organization of this thesis will be introduced at the end of this chapter.

1.1 Background

We live in an age of information explosion. Figure 1-1 appeared in Washington Post at February 11, 2011, which shows how information has exploded over the past two decades. Till 2007, including digital and non-digital data is more than 294.98 billion gigabytes, but this value is only 2.64 billion bytes at 1986. Today, our work, even our life involves a great deal of information. Along with the development of Internet, more and more information existed in traditional medium (e.g. book, vinyl record) was digitized and shared through Internet. Using Internet, we can seek information by search engines. But it is difficult to retrieval our wanted information in enormous search results. The indexes of search engine Google reached 1 trillion URLs at 2008 [2]. Face to the enormous Web resources, about 30% time of intellectual work is spent on retrieval [3]. The experience of using search engine tells us that we cannot retrieve and perceive all search results, because in most cases, these search results are too big to be processed in time. A piece of BBC news represents that most people using a search engine expect to find their wanted information on the first page of results, and stop at the third page [4]. This news indicates that the search results after the page three are almost useless, although search engines can provide a large numbers of search results to people. Therefore, efficient using of Web resources is an important issue for more and more researchers.

Except search engines based on the technique of index data structure, information recommendation is another direction for improving the results of information seeking. Unlike search engines, recommender systems aim to support users in their decision-making while interacting with large information spaces [5]. They recommended items of interest to users based on information that was supplied by users explicitly or collected from users implicitly. Therefore, such systems can

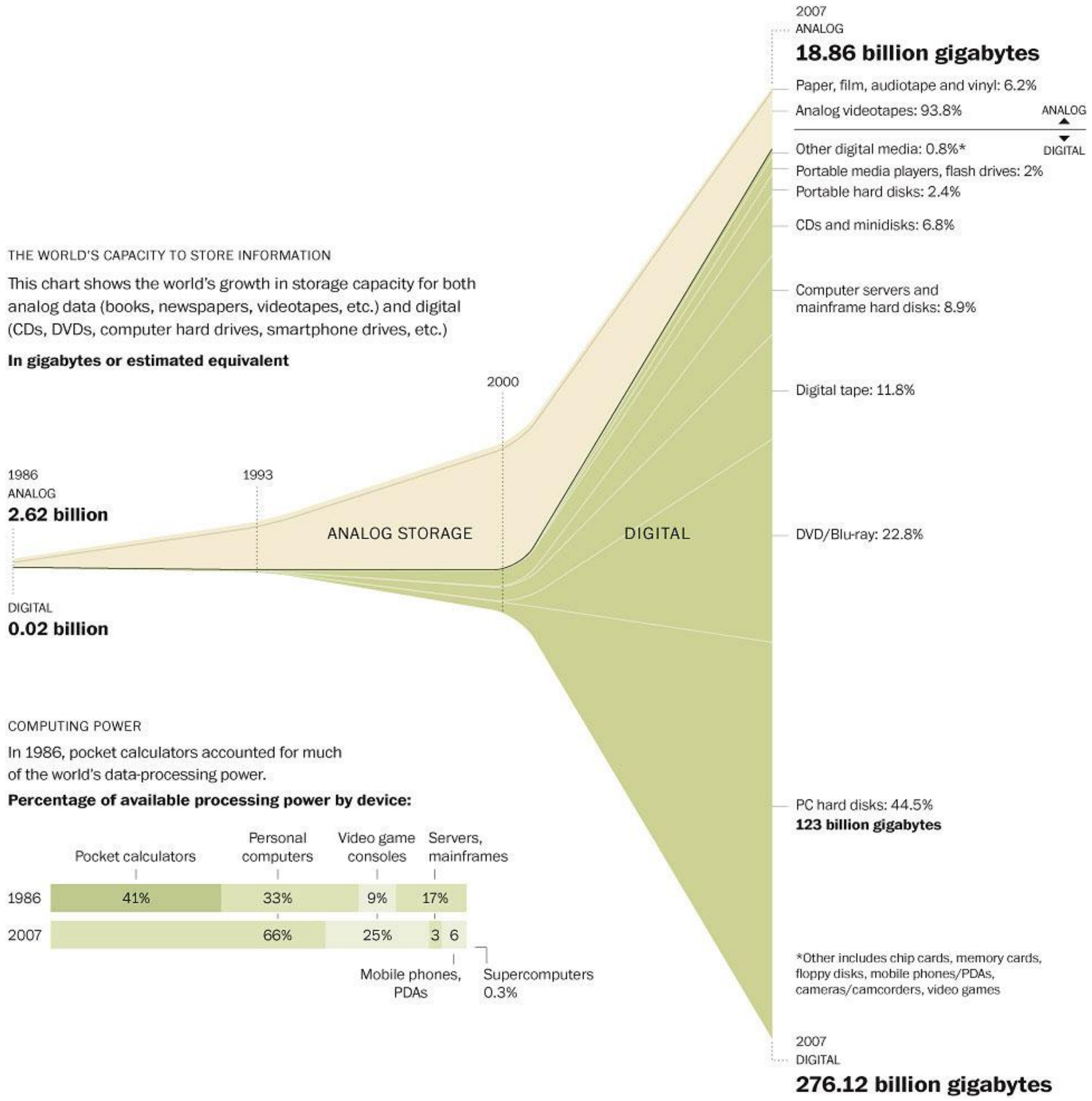

Figure 1-1 The world's growth of both digital data and non-digital data

Washington Post on February 11, 2011 [6]

support users to improve the utilization efficiency of Web resources.

1.2 Purpose of this Study

Information seeking is also regarded as a process that consists of a series of seeking activities. Emerging technologies and paradigms could provide means to realize interactive recommendation systems and support environments that assist personalized information recommendation. Many studies indicate that different action processes led to different results [7] [8]. According to these results, it is possible to advance the efficiency of information recommendation through sharing successful information seeking process based on similar behavior patterns of users.

In this study, an information seeking process represents a purposeful seeking activity to obtain information from Web resource. The information seeking activity consists of a sequence of seeking actions. Through analyzing the seeking process log, content profile and user profile are created and used to describe the relations among users, contents (Web resources) and keywords that are used by users. At the same time, the successful seeking processes are extracted from seeking processes according to implicit feedback of users.

Based on the similarity of user access behavior patterns, we extract a reference user group for a specific target user. After the target user doing a seeking action, a dynamic Bayesian network of seeking actions is created, in which an algorithm of gradual adaptation is used to calculate the posterior probability of the next seeking action from the previous successful information seeking process of the reference user group. The actions with a high posterior probability will be recommended to the target user.

The above approach is called as *Optimized Process Recommendation* model. In this OPR model, we define a set of measures to describe the patterns of users' behaviors of information access, and then an optimized process recommendation algorithm is used to extract and optimize the information seeking process for target users from their reference user group based on the similarity of behavior patterns.

1.3 Organization of the Thesis

This Thesis is organized in six chapters. In Chapter 1, the problem of information overload is posed at first. Around this problem, the merit and demerit of some approaches in improving the information overload is discussed briefly at second. At the end of the this chapter, a new solution of improving the information seeking effect by sharing and optimizing information seeking process is proposed.

After introducing the characteristic of information and some typical approaches of information recommendation, some related works of information access behavior are discussed in the chapter of Related Work.

In Chapter 3, the principle of gradual adaptation model for information recommendation will be introduced. This model is used to track users' focus of interests and its transition by analyzing their information access behaviors, and recommend appropriate information.

User-centric recommendation based on users' behavior analysis is introduced in Chapter 4. Based on the activity theory, the relations of users' seeking behaviors and goal are extracted from seeking process log. It means that the relation between users and contents can be extracted in seeking process log. Furthermore, the relations between user and user are not only considered, but the relations between content and content are also considered by analyzing the similarity of users and contents respectively.

Chapter 5 is used to introduce an application based on our proposed user-centric recommendation approach for individualized learning. In this application, we propose an integrated adaptive framework to support and facilitate individualized learning through sharing the successful process of learning activities based on similar learning patterns in the ubiquitous learning environments. This framework is based on a dynamic Bayesian network that gradually adapts to a target student's needs and information access behaviors. By analyzing the log data of learning activities and extracting students' learning patterns, our analysis results show that most of students often use their preferred learning patterns in their learning activities, and the learning achievement is affected by the learning process. Based on these findings, we try to optimize the process of learning activities using the extracted learning patterns, infer the learning goal of target students, and provide a goal-driven navigation of individualized learning process according to the similarity of the extracted learning patterns.

Finally, the conclusion and future works are introduced in Chapter 6. After giving a summary of this study, the major contribution will be discussed. The last is further works of how to improve the proposal.

2 Related Work

This chapter starts from an introduction about characteristics of information. Based on these characteristics, many researchers pay attention to information recommendation or information access behavior analysis. Such as these related work will discussed in this chapter.

2.1 Characteristic of Information

Alone the human evolution, the communication became important. People study or share knowledge through exchanging information. Generally, information is regarded as a sequence of symbols that can be used to transmit message between people and people, machine and machine or people and machine.

The definition of information was represented by Floridi: information is an objective entity, which can be generated or carried by messages (e.g. words, sentences) or by other products of interpreters; information can be encoded and transmitted, but the information would exist independently of its encoding or transmission [9]. Boell et al. described the attributes of information as six layers – physical world, empiric, syntactic, semantic, pragmatic and social world [10]. By integrating the study of Boell et al., information is considered owning the characteristics as follow:

- Incompleteness: It means that people cannot apperceive the whole aspects of information, because the perception of people is restricted by their owned means.
- Universality: Information exists in cognition, which means information exists in an environment of intelligence.
- Objectivity: Information is an objective reflection of reality.
- Dynamic: The transition of reality leads to the corresponding transition of information.
- Timeliness: Time is a factor of information. A history researcher maybe focuses on the preterit information.
- Depend on medium: Specific information depend on corresponding medium.
- Transitivity: Information can be transmitted by corresponding medium.
- Sharing: Information is shareable, but the sharing base on some specific premises.

The above characteristics of information indicate that information can be transmitted by corresponding medium, and shared each other. More importantly, the characteristics indicate that the importance of the same information may be different at the different time or different situation for people. Therefore, it is an expectation that recommending the appropriate information according to the situation of target users.

2.2 Approach for Recommendation

Based on how recommendations are made, recommender systems are usually classified into three categories – content-based recommendations, collaborative recommendations and hybrid approaches [11] [12] [13]. Balabanovic et al. summed up the three categories as the follows [14]:

- Content-based: the recommended items similar to the ones the user preferred in the past;
- Collaborative: the recommended items were used by other people whose tastes and preferences of the past similar to the target user;
- Hybrid approaches: these approaches combine content-based and collaborative approaches.

In detail, the main means of information recommendation is data mining. Because most of recommended information comes from Web resource, Web mining approaches have been extensively used for information recommendation, in which the final results are directly recommended to target users. Comparing with Web mining, process mining focuses on recommending a process to target user, not a final result.

For information recommendation, the Web mining [15] [16] [17] [18] [19] techniques have been extensively used by researchers. Corresponding to the above categories of recommendations, the web mining has been divided into three main areas: Web Content Mining (WCM) [20] [21], Web Structure Mining (WSM) [22], and Web Usage Mining (WUM) [15] [23] [24]. As an additional area, Semantic Web Mining (SWM) [24] was proposed. Such as the following data types are found in the web and mined in these areas.

- Content data: The text and multimedia in web pages. It is the real data that is designed and provided to users of a web site.
- Structure data: The data consist of organization inside a web page, internal and external links, and the web site hierarchy.
- Usage data: The web site access logs data.
- User profile: The information data of users. It includes both of data provided by users and data created by the web site.
- Semantic data: The data describe the structure and definition of semantic web site.

Although web mining is divided into the four areas, they are associated mostly each other, not exclusively.

We recognize the importance of the content-based web mining such as the WCM and WSM, and pay more attention to the WUM. Because the analysis targets of the WUM and process mining are the same as the web access logs. In the WUM, a new document representation model [24] was proposed recently. This model was based on implicit user feedback to achieve better results in organizing web documents, such as clustering and labeling, which was experimented on a web site with a small number of vocabulary and specific to certain topics. Identifying Relevant Websites from User Activity [25] was another attempt of organizing web pages. It was also based on implicit user feedback but faced the following problem – to improve retrieval accuracy. This model needed to spend more time to train the system.

However, using implicit user feedback has such a problem: although there is a relation between the user implicit feedbacks, there is also a possibility that the uncertainty of implicit user feedback can impair the relation between the web documents clicked by users. Despite this problem, the mining of implicit user feedback can be used to detect the transition of users' focus of interest, and enables us to realize personalized information recommendation. We also noticed that due to the explosive growth of information on the web, web personalization has obtained great momentum both in the research and commercial areas [26]. This encourages us to use implicit user feedback to personalize information recommendation.

Dynamic Link Generation [27] was one of early WUM systems. It consisted of off-line and on-line modules. In the off-line module, the pre-processor extracted information from user access logs and generated records, and then clustered the records to categories. The on-line module was used to classify user session records and identify the top matching categories, and then returned the links that belonged to the identified categories to the user. SUGGEST 3.0 [28] was another kind of WUM systems, but it had only the on-line component. In SUGGEST3.0, the off-line job, like that in Dynamic Link Generation, was realized in the on-line component dynamically. The aim of SUGGEST 3.0 was to manage large web sites. But the size of access logs used to evaluate the system was small and limited. LinkSelector [29] was a web mining approach focusing on structure and usage, in which hyperlinks-structural relationships were extracted from existing web sites and theirs access logs. Based on the relationships, a group of hyperlinks was given to users. Using a heuristic approach, users could access the group to find the information they want.

From the related studies we discussed here, we can see that implicit user feedback is often used in recommendation systems. Further, recommender systems which consist of both off-line and on-line modules have higher performance than those which only have an on-line module. Moreover, concept grouping is more user-friendly because it is easier to retrieve information from a concept group than from an unstructured and not interrelated pool of information.

In the information search field, Pyshkin et al. developed approaches used by web search tools to communicate with users, in order to improve the search quality for various types of information using textual queries, tag-focused navigation, hyperlink navigation, visual features, etc. [30]. Comparing to the traditional keyword search, Klyuev et al. applied the functional approach to represent English text documents, which can take into account semantic relations between words when indexing documents and return only highly relevant documents [31]. Klyev et al. further proposed a strategy that employed Japanese WordNet for the search query expansion, in order to make the search process more convenient for the beginners [32]. Moreover, Shtykh et al. developed a human-centric integrated approach for Web information search sharing incorporating the important user-centric elements and recommendation, namely a user's individual context and 'social' factor realized with collaborative contributions and co-evaluations, into Web information search [33].

Process mining is not a new idea [34] [35] [36], and its applications can be found in the research of business and software development process managements. Many process left their “footprints” in transactional information systems, and their aim is to improve this by providing techniques and tools for discovering process, control, data, organizational and social structures from event logs [37].

Generally, the techniques of process mining have been widely used to discover and analyze business processes based on raw event data, in which there are three types: discovery, conformance and enhancement [38]. The major feature of process mining is to use the event logs to discover, monitor and improve processes based on facts. On the other hand, as an integrative solution, process mining can be considered as an approach between computational intelligence and data mining, and process modeling and analysis [39].

In detail, process mining can be used to extract information about process from event logs. The event was described as the follows [40]:

- Each event refers to an activity (i.e., a well-defined step in the process),
- Each event refers to a case (i.e., a process instance),
- Events are totally ordered.

Furthermore, the process mining was put into context. Such as instance context, process context, social context and external context were used to extract more insights from event data [38].

Although, many researches on process mining focus on how to model a process of business or software development management system, the results of process mining can also be used to improve the recommendation of information seeking process. Because the knowledge will be apperceived in seeking process by users, a recognizing process usually accompanies a seeking process. On the other hand, it means that it may be better to recommend a suitable seeking process than to recommend a final result to user in seeking process.

2.3 Behavior Pattern for Recommendation

Style is the abstraction of pattern. Liu et al. pointed out that students of different learning style chose different learning contents, and their learning styles affected the learning process [41]. Some of the researches are based on the Felder–Silverman learning dimension model that comes from psychology and consists of five pairs of preferred learning style: sensory/intuitive, visual/auditory, inductive/deductive, active/reflective, and sequential/global [42]. Grafter et al. thought a more accurate and detailed description of the five dimensions could be used to improve personalized learning [43] [44]. The study of Halstead et al. shows that the difference of their learning styles between engineering students are small and do not appear to depend on the level of study [45]. The finding of Huang et al. indicates that there is little relation between the assignment score and the learning style score [46].These researches show that although Felder-Silverman learning dimension model could not describe learning style adequately, the following points are true and important

1) There exist different learning styles among different students;

2) Some of the students have similar learning styles;

3) Students of different learning styles whose learning processes are different.

2.4 Summary

In this chapter, we introduced the related works about information recommendation and information access behavior analysis. As described above, collaborative filtering and content-based filtering have respective problems. The commonality is considered by the former one, but is neglected by the latter. The individuality is considered by the latter, but is neglected by the former one. Therefore, to integrate the two approaches is a logical choice. In detail means of recommendation, it was realized that the recommendation based on process is compatible with cognitive process of people. All of these related works orientate our approach towards a process recommendation, in which individuality and collaboration are considered together.

At the end of this chapter, we discussed some studies about information access behavior. These researches of behavior style show that although Felder-Silverman learning dimension model could not describe learning style adequately, we can realize that there exist different behavior patterns for different users, and a different action process could lead to a different result. According to a specific successful user group, the next goal can be inferred for the target users. Continuously, based on the successful process, fruitless actions can be avoided so that the efficiency of information seeking can be improved.

3 Gradual Adaptation Model

Gradual Adaptation model bases on full Bayesian estimation that is used to calculate the posterior probability. After full Bayesian estimation, we introduce the concept and definition of gradual adaptation model. At the end, a simulation is used to evaluate the proposed model.

3.1 Overview

As described above, detecting the transition of users' information access behavior immediately is the key of how to recommend the suitable information. We know that nothing is fixed in the world, everything is changing. Therefore, the needs of people are also changing with their situation. To detect and adapt the transition of user needs, and provide an appropriate recommendation that is compatible with corresponding situation of target users, which is the feature of Gradual Adaptation model (Figure 3-1). The major features of this model are described as follows.

- We divide users' interests into three terms of short, medium, long periods, and by remarkable, exceptional categories - the first one pays a great attention to users' access behavior at current moment, and the second one focuses on haphazard user access.

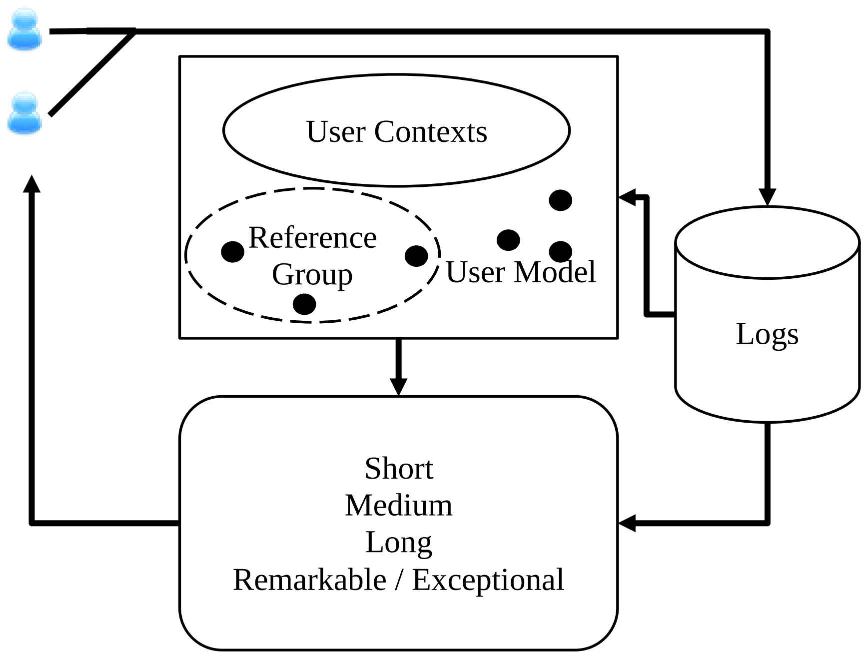

Figure 3-1. Conceptual Chart of Gradual Adaptation Model

- This model is an adaptive one. It can adapt to a transition of users' information access behaviors.
- In the model, training is not needed.

Before introducing the detail of GA model, we will introduce the Full Bayesian Estimation that is the base of GA model.

3.2 Full Bayesian Estimation

In this study, we use Full Bayesian Estimation that has the learning function for the proposed GA model.

The proposed model analyzes the selected link of web pages, and estimates which concept class it belongs to. One link selection is one data sample. The data sample belongs to each concept class. This is expressed as the Equation (3.1).

$$\mathrm{D} = \{D_1, D_2, ..., D_n\} \qquad (3.1)$$

where D_i (i = 1, 2, …, n) represents an aggregate of access samples of concept class that D consists of. D_1 is an aggregate of access samples of concept class D_1. D_2, ..., D_m are the same as D_1 . They are the aggregate of access samples, and belong to concept classes D_2, …, D_m respectively.

We define data sample that is used in Full Bayesian Estimation as follows. If a link that belongs to a concept class D_m is clicked, we use d_t to describe the number of click times of D_m, and d_f to describe the number of click times that concept class D_m is not clicked (i.e. other concept classes are clicked). For the history logs (not including current day), we use a variable a_t to describe the number of click times of concept class D_m, and a_f to describe the number of click times that concept class D_m is not clicked.

For example, if the whole click times is 6 at current day, and the 2 times belong to concept D_m , it means $d_t = 2$, and $d_f = 4$, then we can accord (2) [47], and gain the click probability of the concept D_m is $2/6$.

$$\theta^* = \frac{d_t}{d_t + d_f} = \frac{d_t}{d} \qquad (3.2)$$

Equation (3.2) is called as Maximum Likelihood Estimation. Because the empirical value is disregarded by Maximum Likelihood Estimation, haphazardness can give a big influence on the estimation result.

But in Full Bayesian Estimation, the join of Prior Distribution (based on the history click samples) and Likelihood Estimation is used to calculate the Posterior Distribution Θ. Its expression is described as follows:

$$\begin{aligned} P(D_{m+1} = t \mid Đ) &= \int P(D_{m+1} = t, \theta \mid Đ) d\theta \\ &= \int P(D_{m+1} = t \mid \theta, Đ)\, p(\theta \mid Đ) d\theta \\ &= \int \theta p(\theta \mid Đ) d\theta \qquad (3.3) \end{aligned}$$

where Đ is a data collection which consists of (D_1, D_2, …, D_m), and is used to describe the Likelihood Estimation. The integral calculation of Full Bayesian Estimation as show in Equation (3.3) is very complicated. Generally, it needs the following premises to make it calculable.

- Each sample in Đ is independent with each other, and satisfies *iid* (independent and identically distributed) assumption;
- About the current click times d_t and d_f, theirs prior distribution satisfies Bate Distribution $B[a_t, a_f]$.

Thus, the Full Bayesian Estimation equation can be expressed as follows [47]:

$$P(D_{m+1} = t \mid Đ) = \int \theta p(\theta \mid Đ) d\theta$$

$$= \frac{\Gamma(d_t + \alpha_t + d_f + \alpha_f)}{\Gamma(d_t + \alpha_t)\Gamma(d_f + \alpha_f)} \int \theta \theta^{d_t + \alpha_t - 1} (1-\theta)^{d_f + \alpha_f - 1} d\theta$$

$$= \frac{d_t + \alpha_t}{d_t + d_f + \alpha_t + \alpha_f} \tag{3.4}$$

According to Equation (3.4), if the number of the current samples is small, prior distribution has a big contribution on the result. On the contrary, if the number of the current samples is big, prior distribution has a little contribution on the result.

3.3 Concept and Definition

In this study, we try to use dynamic sampling of user information access behaviors, instead of the static period. Specifically, we define a top probability group that consists of a number of concept classes which have high posterior probability. When the probability sequence of the group changed, a new sampling point is set. We call it the sampling point of sequence change. When a change in the member of group is observed, another new sampling point is set. We call it the sampling point of member change. A period that is between the two adjacent sampling points of sequence change is defined as the short period. Similarly, a period that is between the two adjacent sampling points of member change is defined as the medium period. By this, we can get the concepts representing users' focus of interests for each period of their Web activities, similarly as done in [48]. In addition to the short and medium periods, we also define the long period to describe a user's long-term focus of interests. We assume it is the full access period of users or the maximum period allowed by privacy laws related to collecting personal information about users on the Web.

In this study, we set two dynamical periods – short and medium. A static long

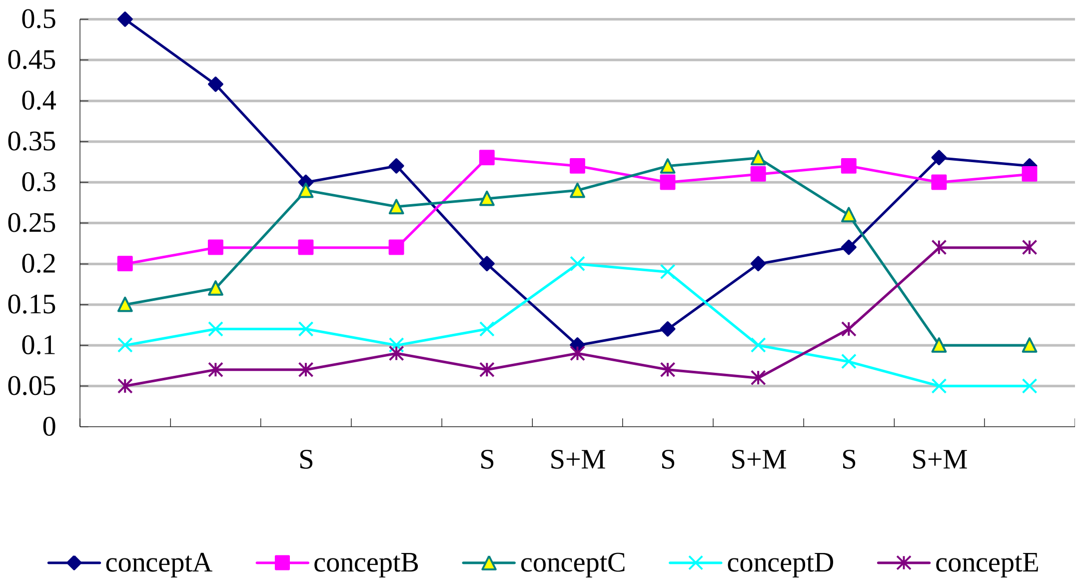

Figure 3-2. Period Definition

period is set as a fixed one, for example, 30 days. In the proposed system, users' Web access information is collected by a unit of one HTTP session. If there are samples that are from two sessions for a concept class, its initial probability can be estimated.

Due to the introduction of dynamic sampling in the above, we assume the top group number, for example, is 3. It means there are three concept classes that belong to the top reuse probability group. As shown in Figure 3-2, the sequence of top three concept classes group changed at the point "S", thus, "S" is a sampling point of short period, and the period that is between two adjacent "S" is a short period. At the point "S+M", the members of top three concept classes group change. Therefore, "S+M" belongs to the sampling point of both short and medium period. A period that is between two adjacent "S+M" is a medium period. Short period is designed to reflect temporary interests of users. Medium period is designed for an interest that is affected by some factors, i.e., this interest is relatively stable during a period. Long period is designed for a long-term user interests. In the next section, the different features of these three periods will be shown. Besides the three periods, remarkable

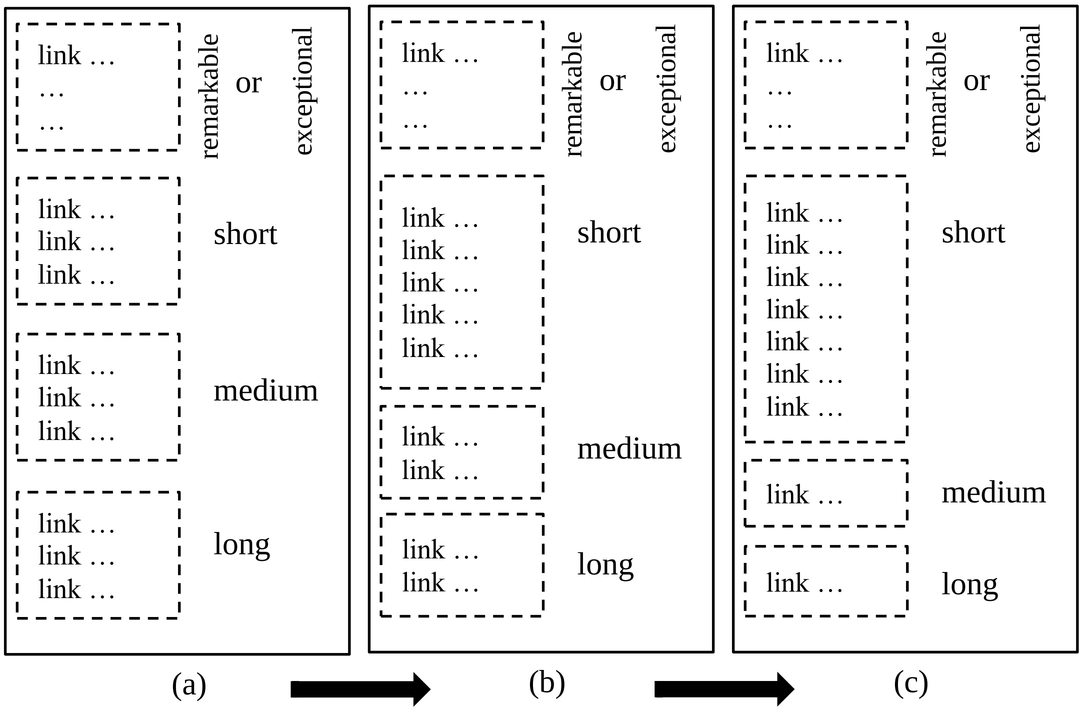

Figure 3-3. The Transition of Gradual Adaptation Model

category means high degree of a user's focus of interests in a particular concept class. Exceptional category is an aggregate of a concept class that has a little chance to be clicked by users, but may be useful for users in the future.

When received a search query from user, it will be checked if there is a remarkable concept class of the user. As shown in Figure 3-3, if remarkable concept class exists, the corresponding links of remarkable concept class will be put at the top of a recommendation page, and then, a certain number of links from each period will be put below the remarkable links respectively. Of course, these links belong to the concept class which has high probability in each period.

If a remarkable concept class is not found, it will be checked if an exceptional concept class exists. If an exceptional concept class exists, the corresponding links of the exceptional concept class will be extracted in a random manner and put at the top of a recommendation page. The same as in the previous case, these links belong to the concept class which has high probability in each period. If both remarkable and exceptional concept classes do not exist, the same number of links will be extracted from each period respectively. Using the described approach, a hint about which concept class is their hot concept class or which concept class is the concept class they almost forgot is given to target users.

Figure 3-3 (a) is the first response to a user. If the user makes a decision on a link and click it, the concept class, keyword and period or category information about the link will be regarded as an implicit feedback. Obtaining such information, user's demands will be inferred through GA model.

As show in Figure 3-3 (b), if the selected link belongs to short period, the links number of short period will be increased. At the same way, the links number of other terms will be reduced.

As show in Figure 3-3 (c), if the link of short period is clicked continuously, the links number of short period will be increased to a maximum number, and the number of other links of each period will be reduced to a minimum number. If a link that belongs to the short period is not clicked continuously, and another link that belongs to the other period is clicked - for instance, a link that belongs to the long period is clicked - in this case, the number of recommended links for the short period will be reduced, and at the same time, the number of links for the long period will be increased.

The same things occur in case of other periods, and GA model can detect the change and redress the recommendation result. Therefore, GA model can give a high satisfaction rating to users.

3.4 Simulation and Evaluation

In order to verify the operability of the proposed GA model, we pre-produced the model. The system was built by open source software: Java, Tomcat, MySQL, and Nekohtml were used.

For the simulation, we consider three basic cases to evaluate the system. The first case is a user who has a long-term interest. In this case, the probability of the interested concept class ought to keep a high rate in long period.

The second case is a user who has a temporary interest. The user access the concept class of temporary interest sometime. In this case, this concept class ought to keep a low rate in the three periods.

The third case is a user who has two interests, and these interests are affected by some factors easily. In the case, there is a possibility that the probability of the relation concept class can change hugely in the short or medium period, but not in the long period.

3.4.1 Creating of Concept Class Base

We use Wikipedia on DVD Version 0.5 [3] (we refer to it as Wikipedia 0.5 for brevity) as the test data. The lowest categories in Wikipedia 0.5's topic hierarchies are used as the concept classes in the simulation. Based on Wikipedia 0.5, we gained more than 2000 web pages that belong to 180 concept classes. These concept classes were ready in advance, and saved in the Concept Class Base.

[3] http://www.wikipediaondvd.com/nav/art/d/w.html

3.4.2 Setting of Simulation Case

At first, we define a concept group that consists of top three probability concept class. For case one, the concept class “Religion Divinities” is assumed to be used almost every session, and the assumed number of clicks (a user’s accesses) is set to 0 and 16 per access session. It means the user is interested in the concept class, and has a long-term interest in the concept class.

For case two, the concept class “Philosophical thought movements” is assumed to be used per two sessions, and the assumed number of clicks was set to 0 and 5 at first half of test period. But the click number becomes more than 15 at the second half of test period. It means the user has little interest in this concept class at first, but becomes more and more interested in it.

For case three, two concept classes of “Arts Museums galleries” and “Artists” are assumed to be used, and the number of clicks is dynamically varying. Most times it is set to 0 to 8, but sometimes it is set to 0 to 15 (likes case two), some other times it is set to 0 to 20 (large than case one).

Obviously, concept classes “Religion Divinities” and “Philosophical thought movements”, and concept classes “Arts Museums galleries” and “Artists” are similar respectively in the cases described above. It means that similar concept classes have similar keywords. We expect that our model can differentiate the concept classes even if they contain similar keywords, and gain the results as we explained above.

3.4.3 Simulation Results

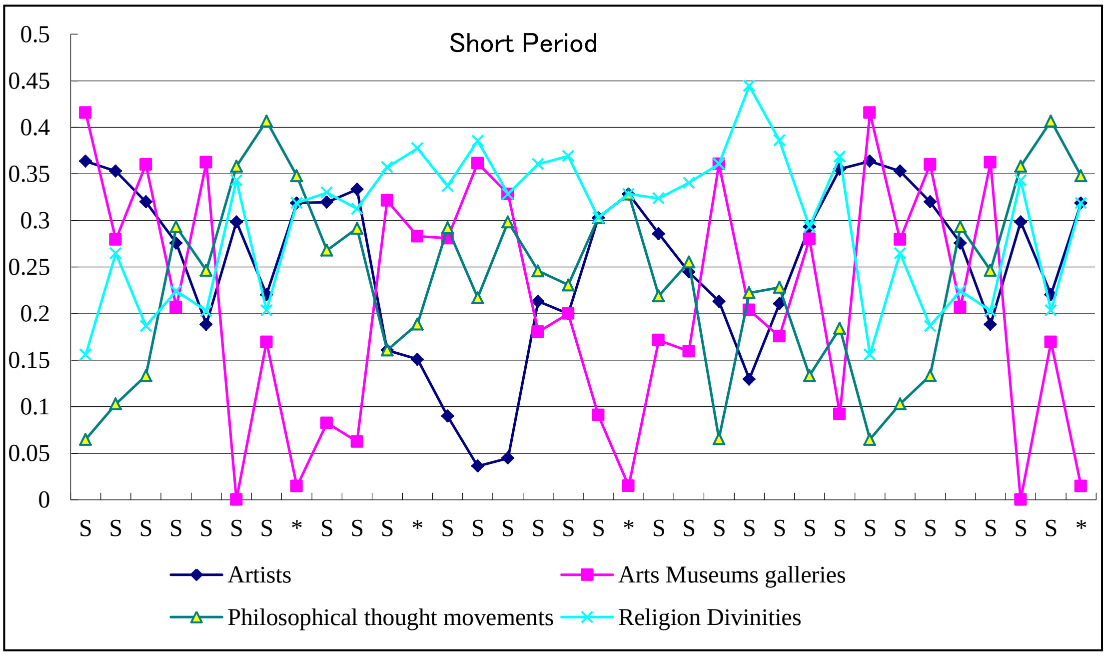

Figure 3-5. Probability of Concept Classes in Short Period

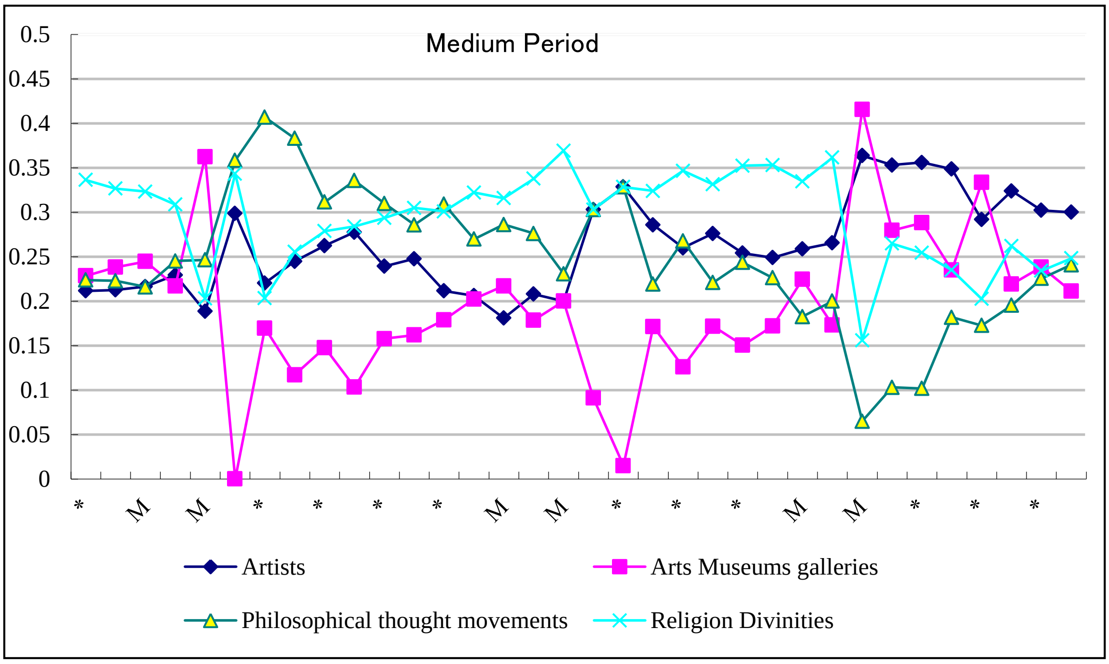

Figure 3-4. Probability of Concept Classes in Medium Period

We simulated the three test cases during a period of 60 access sessions, and obtained the results. The results are what we expected, showing high adjustability and adaptability of the proposed model.

In short period, we can see the movement of the concept rate changing frequently,

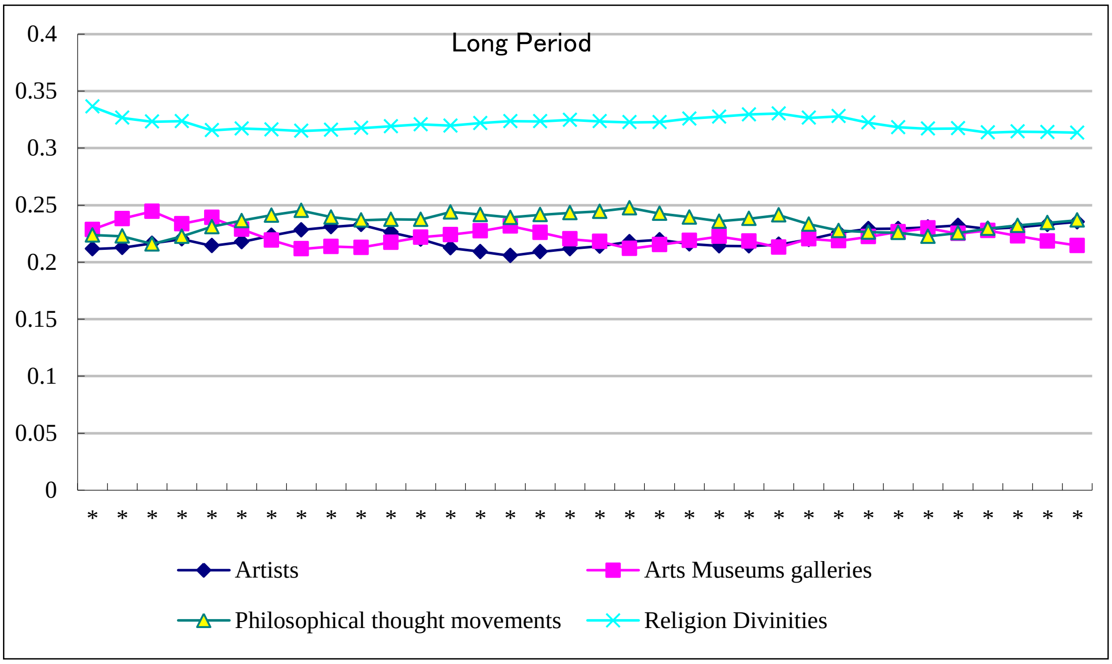

Figure 3-6. Probability of Concept Classes in Long Period

and its frequency is higher than in the static period [1]. It means that dynamic sampling can detect and track smaller transition of users' focus of interests than the static period (Figure 3-4). It results in line with our expectations that concept classes of high changeability represent the user's short-term focus of interests.

In medium period, the change is becoming smaller. But the probability of concept classes in Cases two and three are bigger than in case one in some points (Figure 3-5). The system regards a concept class which concept change rate is smaller than that of short period as the medium-period focus of interests of the user.

In long period, the change becomes even less dramatic. As shown in Figure 3-6, the long-term interests of user correspond to the "Religion Divinities".

In order to identify the change rate of three periods, a change rate threshold needs to be determined by more experimentation. Then, the system can identify the three periods more accurately and recommend appropriate concept classes to users.

From the simulation results, we found that the proposed model adapts well to the transition of user's focus of interests as we expected. Thus, if a concept class is used frequently, it ought to have a high probability in long period. If the concept class is used to a certain extent, it ought to have a quick change in short or medium period. If the concept class is used rarely, the rate ought to keep at a low level.

Specially, comparing with the results of the static period [1], we verified that dynamic period is capable of detecting and tracking a small transition of users' focus of interests, and adapting to the individuality of users. The results demonstrate that the proposed model is operable and effective for modeling situations similar to those in the above-mentioned cases.

3.5 Summary

In the above, we have introduced a GA model for estimation of user information access behavior, based on Full Bayesian Estimation with a learning function, in order to solve the uncertainty problem caused by differences in user information access behaviors. A variety of users' information access data are collected and analyzed in terms of short, medium, long periods, and by remarkable and exceptional categories. We have further established the proposed model with three assumed cases to show operability and effectiveness of the model with experimental simulation results.

The simulation results have shown that the proposed model can recognize the transition of users' access behaviors (web page selections, in particular) sensitively in short period. The users' long-term interest kept a high probability in long period. The three periods of GA model can correctly distinguish long-term and temporary interest of users.

4 User-Centric Recommendation

The research of User-Centric is a focus in recent years. Around this focus, the well-known activity theory will be introduced firstly. Based on this theory, an *Optimizing Process Recommendation* model is introduced secondly. The scenario and simulation of *Optimizing Process Recommendation* model are described at the end of this chapter.

4.1 User-Centric Thinking

The concept of user centricity originated from software development. Patton indicated that user centricity isn't just asking user what they want [49]. He suggested understanding users by creating a user model and collaborating effectively with them. Arbanowski et al. described that "centric" means adaptable to user requirements and a certain user environment [50]. Recent years, many studies focus on user-centric approach and use this approach to recommendation [51] [52] [53] [54]. The characteristic of these studies is that the interaction between user and system is considered. Through the interaction, recommender systems adapt to the transition of users, and revise the recommending results in order to make them satisfy the needs of users.

The concept of User-Centric Recommendation is described as shown in Figure 4-1. A recommendation system extracts experience from reference users firstly, and

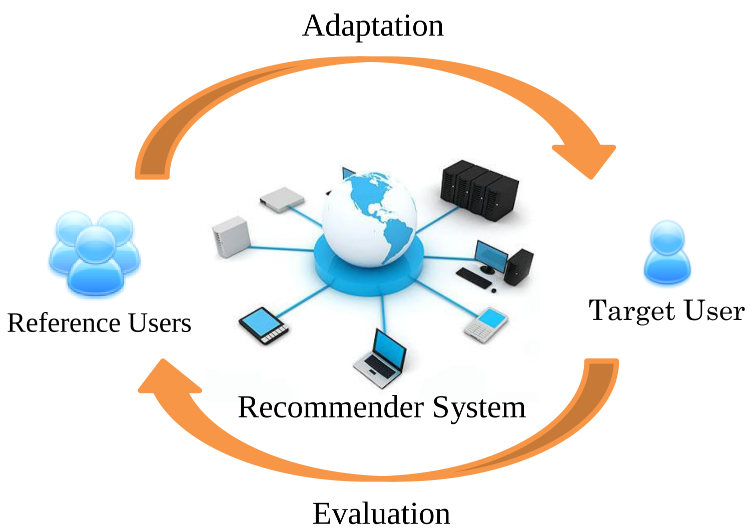

Figure 4-1. The Concept of User-Centric Recommendation

then provides recommendation to target user. The response of target user is regarded as an implicit feedback, and used to evaluate the recommended items. According to the evaluation results, the recommender system revises recommending items. For example, if the evaluation isn't satisfactory, the recommender system can revise recommending items by detecting the transition of target users or rearranging the reference users whose behavior patterns are similar to the target users more.

The detail of user-centric recommendation based on the proposed GA model and user-centric approach will be introduced in next section.

4.2 User Behavior and Activity Theory

According to the well-known activity theory, activities consist of actions or chains of actions, in which actions in turn consist of operations. The activities, actions and operations can be regarded as corresponding to motive, goal and conditions [55]. When conditions change, an operation can return to the level of action.

As shown in Figure 4-2, the activity is regard as a longer term process, which is used to achieve a corresponding motive (i.e., want to search some information about program languages). The action is regarded as a shorter term process, which is used to achieve a corresponding goal (i.e., a step from inputting a set of keywords to the end of checking the corresponding search results). And the current moment status is regarded as an indicator to describe conditions at the moment (i.e., reading a search result, page up or page down). That is, the longer term process of activities consists of a set of actions with a specific sequence, while the shorter term process of actions consists of a set of operations with a specific sequence. Generally, in order to achieve a motive or a goal, users need to complete a longer term process of activity or a shorter term process of action. Therefore, based on these, we can infer a motive or a

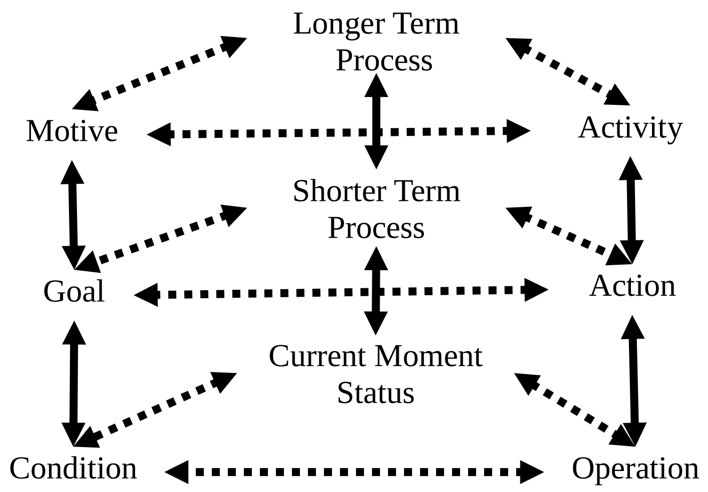

Figure 4-2. Relation of motive, activity and process

goal through analyzing longer term processes of activities or shorter term processes of actions.

As shown in Figure 4-3, the upper case of M, G and A denote Motive, Goal and Action respectively. For each goal, there exists a corresponding action. A set of arrow lines starting at the circle node and ending at the triangle node denote a specific activity. There are a number of different activities that can be used to achieve the same motive, in which some one may suit to some user and the other one may be suit to another user.

Based on some of previous successful activities, we can obtain the actual figure like the conceptual image. Using this figure, we can infer the next action or goal of each current action in the successful activities. But not all of the activities fit to every user. For a target user, we can create such as the in Figure 4-3 from the successful activities which belong to a reference user group whose information access behavior patterns are similar to the target user.

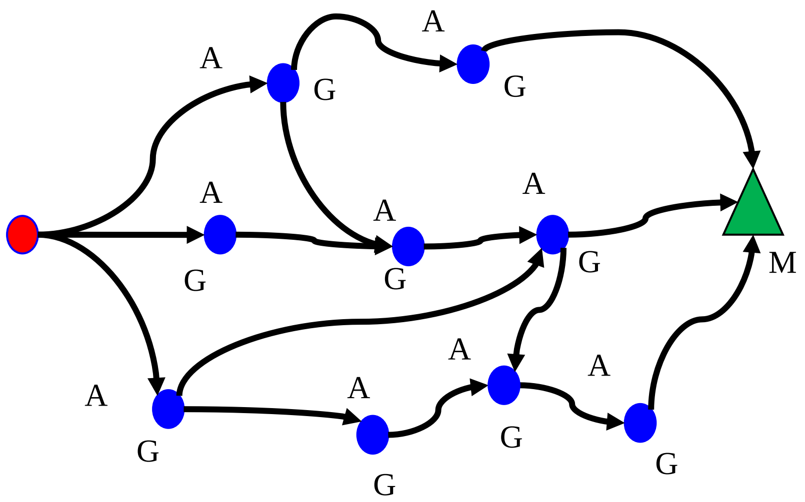

Figure 4-3. Conceptual image of seeking process

For an information seeking process, when a user wants to seek something from web pages, we can say the user has a motive of information seeking. In order to achieve the motive, an information seeking process is started. This seeking process (access activity) is divided into a series of access actions and used to achieve the corresponding seeking goal with a sequence. For example, inputting the keywords and perceiving results are the first access actions. And then, based on the sought results, inputting other keywords and perceiving results again are the second access actions. These access actions will be repeated, till this specific user achieves the motive.

4.3 Optimizing of User Seeking Process

As described above, we know that the motive can be inferred by observing behavior of user, and a different action process could lead to a different result. Therefore, we expect that successful process can help the other users whose information access behavior patterns is similar to the successful users. In order to realize the aim, we propose a mode namely *Optimized Process Recommendation* model [56] that adapts

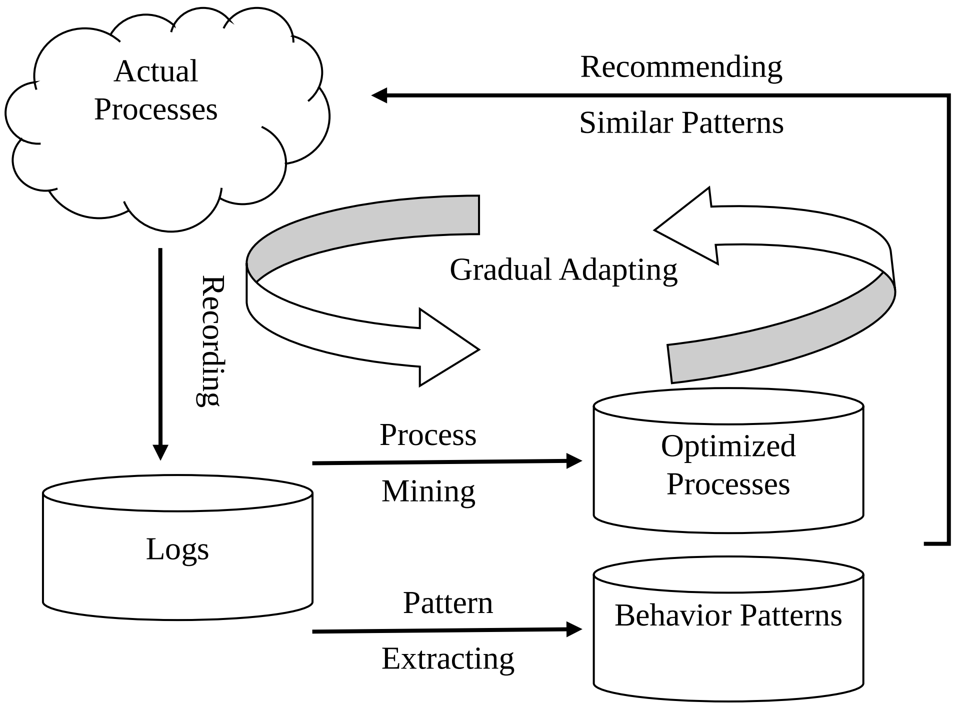

Figure 4-4. Concept of optimized process recommendation model

to a target user's transition of needs and behaviors of information access.

Figure 4-4 shows the concept of optimized process recommendation model. The actual processes denote the processes that were taken by users actually. When users seek something from web pages, their actual process will be recorded into logs. These logs are then used to extract the information access behavior patterns, and also used for process mining. The aim of pattern extracting is to extract information access behavior patterns of users from logs. The aim of process mining is extracting successful seeking experience, and then optimizing the seeking process from the extracted successful seeking process according to the similarity of users' behavior patterns. Finally, the optimized seeking processes are recommended to the target users whose information access behavior patterns similar to the reference user group. Here, the reference user group represents the users who have successful experience in their information seeking processes.

The optimized process recommendation model is based on our previous study called as *Gradual Adaptation* model [1]. Therefore, the use of recommendation results can be monitored and re-analyzed, then the analyzed results are used to advance the recommendation effects gradually. This process is repeated till the target users obtain their satisfying results.

The details of how to extract information access behavior patterns and how to optimize the seeking process will be introduced in the next section.

4.4 User Access Behavior Patterns and Similarity

In this section, to describe the relationships among users, contents of web pages, keywords and tags, firstly, we introduce a set of definitions. After that, based on the description of the seeking process in details, we further give the architecture of a seeking process recommender system based on optimized process recommendation model.

4.4.1 Definitions for User Profile and Content Profile

User profile and content profile are used to represent the feature of users and contents (web pages).

The following elements are used to describe user profile:

- Keyword: used by users
- Web page: accessed by users
- Tag: belong to the web page accessed by users (in this study, we use Hatena Bookmark [4])
- Access time: user access time

[4] http://b.hatena.ne.jp/

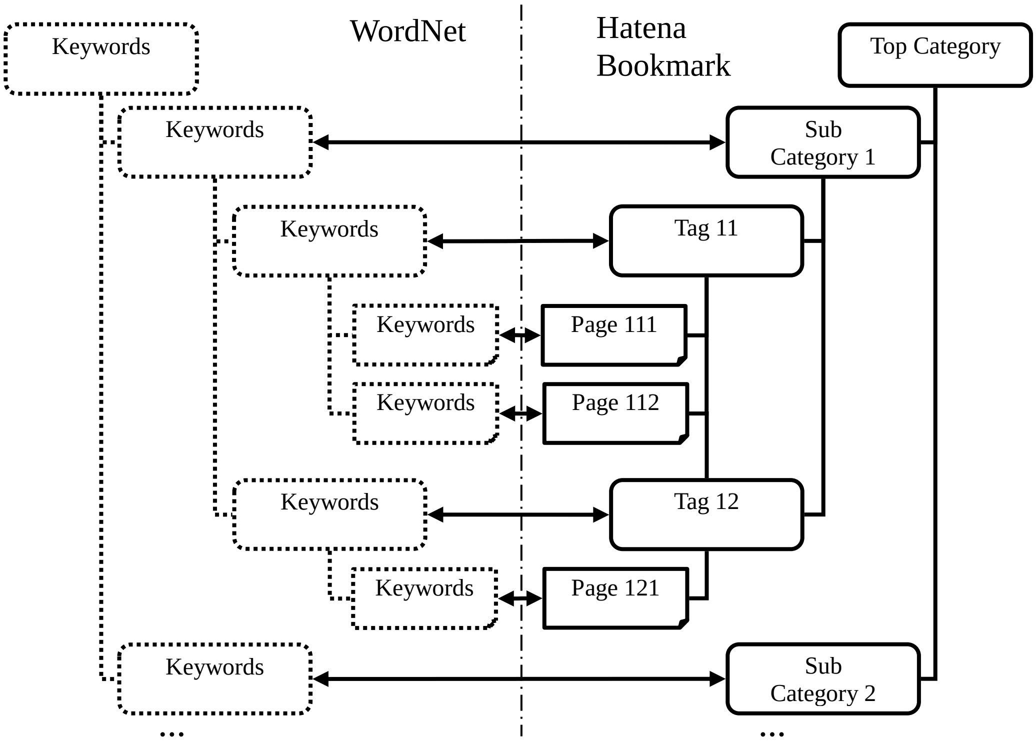

Figure 4-5. Keywords-web page tree

- Access date: user access date
- Location: users' location data

Content profile is based on the keyword-web page tree, and its element is given as follows.

- relation of tag (e.g., Hatena Bookmark) and web page
- relation of tag category (e.g., Hatena Bookmark) and web page
- relation of keyword and web page (in this study, we use WordNet [5])

Figure 4-5 shows the keyword-web page tree, which is based on WordNet and Hatena Bookmark. The basic idea is that if a keyword relates to a web page, it also relates to the Hatena Bookmark tags that the web page belongs to. Furthermore, if a web page includes a keyword, the keyword's thesauruses based on the WordNet also relate to the web page.

[5] http://wordnet.princeton.edu/

4.4.2 Description for Information Seeking Process

In order to measure information seeking process, we define a number of measures such as seeking process starts, seeking process ends and successful seeking process. The details are summarized as follows.

- Seeking process start: When a set of search keywords are inputted and the search button of search engine is pressed down, it means a seeking process is started.
- Seeking process end: The keywords changed and the meaning is different with the previous keywords based on WordNet. This means a seeking process is over, and a new seeking process starts.
- Successful seeking process: At the point of seeking ends, if the used keywords relate to the accessed web page at the previous step based on the content profile, this seeking process is regarded as a successful seeking process.

For example, when a user inputs a set of keywords, such as "programming languages", and typed the enter key, this time point is regarded as the seeking process start. The user may change the search keywords in the seeking process. If the changed search keywords are not the synonym for "programming languages" according to the thesaurus of WordNet, this time point can be regarded as the seeking process end. At the time point of the seeking process ends, if the accessed web page of the previous step relates to the "programming languages", this seeking process can be regarded as a successful seeking process.

4.4.3 Description for Information Access Behavior Patterns

Generally, when a user wants to seek something from web pages, a motive should be generated. For achieving the corresponding motive, the user accesses to the proposed system, and further starts the information seeking process. In details, the user needs to consider some detailed goals for achieving the motive, for example, to consider some good keywords, to visit web pages and estimate if the results are expectant. In order to achieve the goals, the user needs to complete the corresponding action process. If the results disaccord with the goal, the user needs to adjust corresponding goals. Based on the user's seeking process, the system can obtain the user's seeking action logs, such as the keywords used by the user, the web pages accessed by the user. Furthermore, the user's access time, weekday, location can be obtained and recorded into access logs.

Based on the above action logs, the system can extract user's behavior patterns. Equation (4.1) is used to describe users' similarity of information seeking behaviors.

$$
\begin{aligned}
& sim(u,u^{'}) = a_1 sim(k,k^{'}) \\
& + a_2 sim(w,w^{'}) \\
& + a_3 sim(t,t^{'}) \\
& + a_4 sim(p,p^{'})
\end{aligned}
\quad (4.1)
$$

where,

sim(*u,u'*) denotes the similarity of two users *u* and *u'*,

sim(*k,k'*) denotes the similarity pattern of keywords used by users *u* and *u'*,

sim(*w,w'*) denotes the similarity pattern of web pages used by users *u* and *u'*,

sim(*t,t'*) denotes the similarity pattern of Hatena Bookmark Tag which belongs to the web page that has been visited by these two specific users,

sim(*p,p'*) denotes the similarity pattern of the two users' profiles, a_1, a_2, a_3 and a_4 are the weight.

In details, the following equations (4.2)-(4.5) are employed to calculate the similarities discussed above respectively.

$$sim(k,k') = |k \cap k'| / |k \cup k'| \qquad (4.2)$$

where k and k' indicate the keywords used by users u and u'.

$$\begin{aligned} sim(w,w') &= b_1 sim(w_1, w_1') \\ &+ b_2 sim(w_2, w_2') \end{aligned} \qquad (4.3)$$

where

$sim(w_1,w_1')$ indicates the similarity of the web pages that are accessed by users u and u';

$sim(w2,w2')$ indicates the similarity of the web pages that belong to the same Hatena Tag, but the details are different.

$$sim(t,t') = |t \cap t'| / |t \cup t'| \qquad (4.4)$$

where t and t' indicate the Hatena Bookmark tags used by users u and u'.

$$sim(p,p') = p \times p' / |\sqrt{p^2}| \times |\sqrt{p'^2}| \qquad (4.5)$$

where p and p' indicate the specific user data, such as access time, access date and location of users u and u'.

Based on the above equations, a reference user group can be extracted for the target users. And then, the successful seeking processes are extracted from the reference user group. We will discuss these in details in the next section.

4.5 Mechanism for Information Seeking Process Optimizing

In this sub-section, we focus on introducing the algorithm and architecture of optimized process recommendation model.

4.5.1 Optimized Process Recommendation Algorithm

OptimizedProcessRecommending()

Input: User action $o_u = (k, w, t, p)$, where k: keywords; w: web pages; t: tags that accessed web pages belong to; p: user profile.
Output: next action $O'[]$

1. Initialize top tag group
   input o_u;
   if (target user has log) {
   select n sets actions $O_t \supseteq o_u$ from target user's log, the posterior probabilities of O_t are higher than a threshold;
   } else {
   select n sets actions $O_t \supseteq o_u$ from reference user's log, posterior probabilities of O_t are higher than a threshold;
   }
2. Detect the transition of process
   2.1.Calculate the posterior probabilities $\Theta[i,j]$ of current top tag group
   current top tag group T_c;
   for(i = 1; i < number of T ($T \in T_c$); i ++) {
   for (j= 1; j < number of P ($P \in T$); j ++) {

$$\theta[i, j] = \left(\frac{1}{\| T \|} \times \frac{\underset{P_j \in Ti}{P_j}}{\sum_{P_j \in T_i} P_j}\right) / \frac{P_j}{\sum P_j}$$

   , where $\Theta[i,j]$ is the posterior probability of a web page P_j that belongs to tag T_i, and accessed by user, $\|T\|$ is the number of tags that relate to o_u, P is the number of accessed web pages that belong to tag T.
   }
   }
   create current top tag group T_c by $\Theta[i,j]$;
   2.2.Check the transition of top tag group
   top tag group T_t; user focus T_f;
   if (member of T_t == member of T_c) {
   if (sequence of T_t == sequence of T_c) {
   return the $T_f \subseteq T_c$ with top three high posterior probability;
   } else {
   shorter term process changed;
   return the $T_f \subseteq T_c$ with top three high posterior probability;}
   } else {
   longer term process changed;
   return the $T_f \subseteq T_c$ with top three high posterior probability;}
3. Recommend the next action
   for (i = 0; i < 3; i ++) {
   o_n = the next action that belongs to $T_f[i]$ from reference user group;
   $O'[i] = o_n$;
   }
   return O';

Figure 4-6. Optimized Process Recommendation Algorithm

The algorithm shown in エラー! 参照元が見つかりません。 is used to extract and further optimize the successful information seeking process. As discussed above, the seeking process includes a longer term process that consists of a series of actions, and a shorter term process that includes a series of operations. When a target user completes an action, the system will calculate the posterior probability of the next action from the successful seeking process, and then recommend the high posterior probability operation to target user.

4.5.2 Architecture of Optimized Process Recommender System

This study proposes an approach to detect users' information seeking motive or goal, and extract users' behavior patterns through investigating users' information seeking process and situation. Furthermore, by detecting the transition of a target user's information seeking process, a dynamic Bayesian network is generated from a reference user group whose behavior patterns are similar, and then, an optimized information seeking process is recommended to the target users.

The system has two major functions:

(1) creating an adaptive user model by capturing users' information access behavior and interactions with the system and/or other users;

(2) matching and recommending appropriate information and available processes to a target user, and delivering them.

The prototype system consists of several main modules, such as User and Content Profile Creator, Pattern Analyzer/Similarity Marcher, Seeking Process Optimizer, and Gradual Adaption Recommender, which are shown in Figure 4-7.

First of all, the User Profile Creator and Content Profile Creator work as a context base in the system. The User Profile Creator also takes into account of other

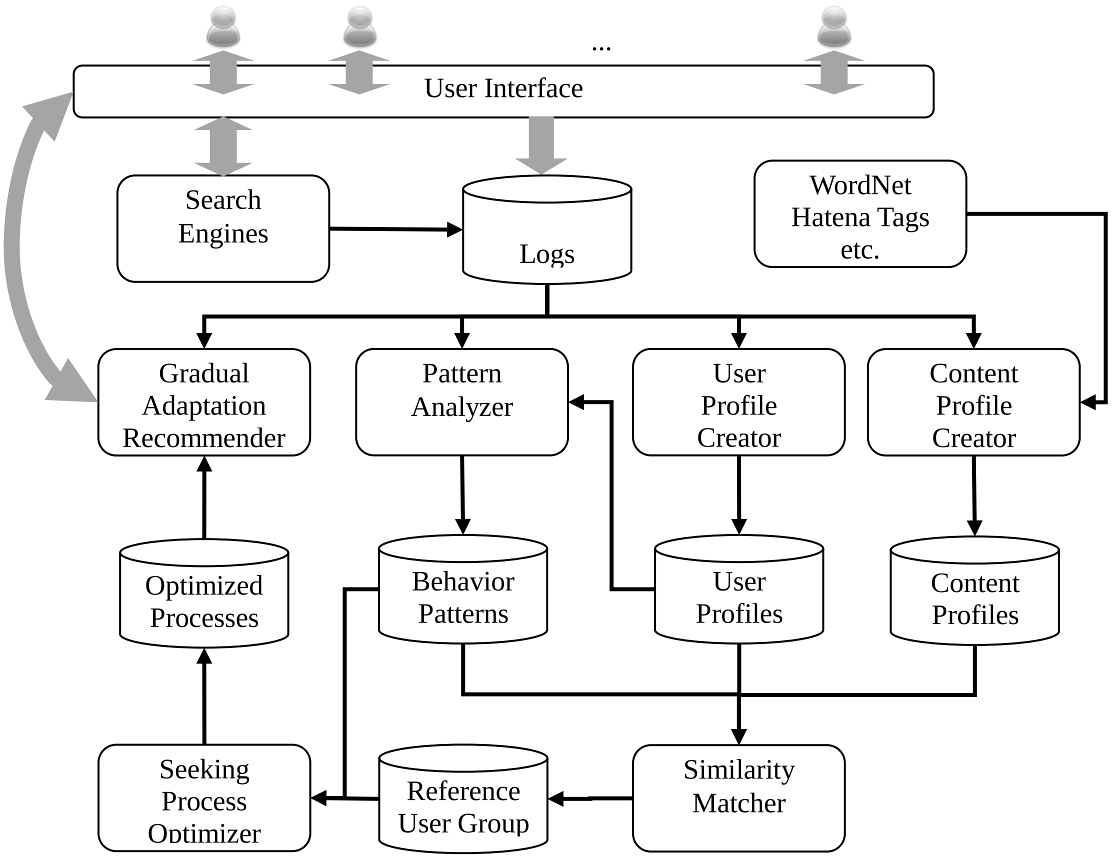

Figure 4-7. Architecture of optimized process recommender

users' behavior and feedback results, which leads to a dynamic group profiling. In our proposed system, users' evaluations on an information seeking process (a sequence of using information seeking) are shared among all users. Content Profile Creator is based on WordNet and Hatena Bookmark tag, and used to extract the relation among Hatena Bookmark tag, web page and keyword.

The Pattern Analyzer and Similarity Marcher are used to extract information access behavior patterns and create a reference user group for target users. The Seeking Process Optimizer is used to optimize the successful seeking process. For a target user, a reference user group is extracted by Similarity Matcher. And then, the successful information seeking processes of the reference user group are analyzed by Seeking Process Optimizer. The analyzed results are recorded into Optimized

Processes.

The Gradual Adaptation Recommender is originally conceived to detect and track information access behaviors (including various information seeking related activities, such as accessing to web resource in the system) and then inference the transition of user's possible current moment status and concerns over a certain time span, such as a shorter or longer term process. Users' interactions with the system and other users in the system, and interactions with the whole web world as well, are monitored and analyzed to help creating a more refined user model. Gradual Adaptation Recommender is designed based on Bayesian Network, which is used to calculate posterior probabilities that are applied for recommending information and predicting users' possible actions and/or potential change of focuses and concerns. Based on the calculated posterior probability, Gradual Adaptation Recommender re-ranks the search results of search engines, and recommends appropriate seeking process and results with possible various metadata (such as efficient seeking sequences) according to the established user model, and navigates them to the user as a web service utilizing the service mash-up approach.

4.6 Scenario and Simulation

In this section, we use a scenario to describe how the proposed system works to support and facilitate information seeking process.

4.6.1 Scenario

In order to introduce the scenario, a flowchart is shown in Figure 4-8. When a user wants to seek some information by the proposed system, the first step is to input a set of keywords. And then, the system will return the results to the user according to

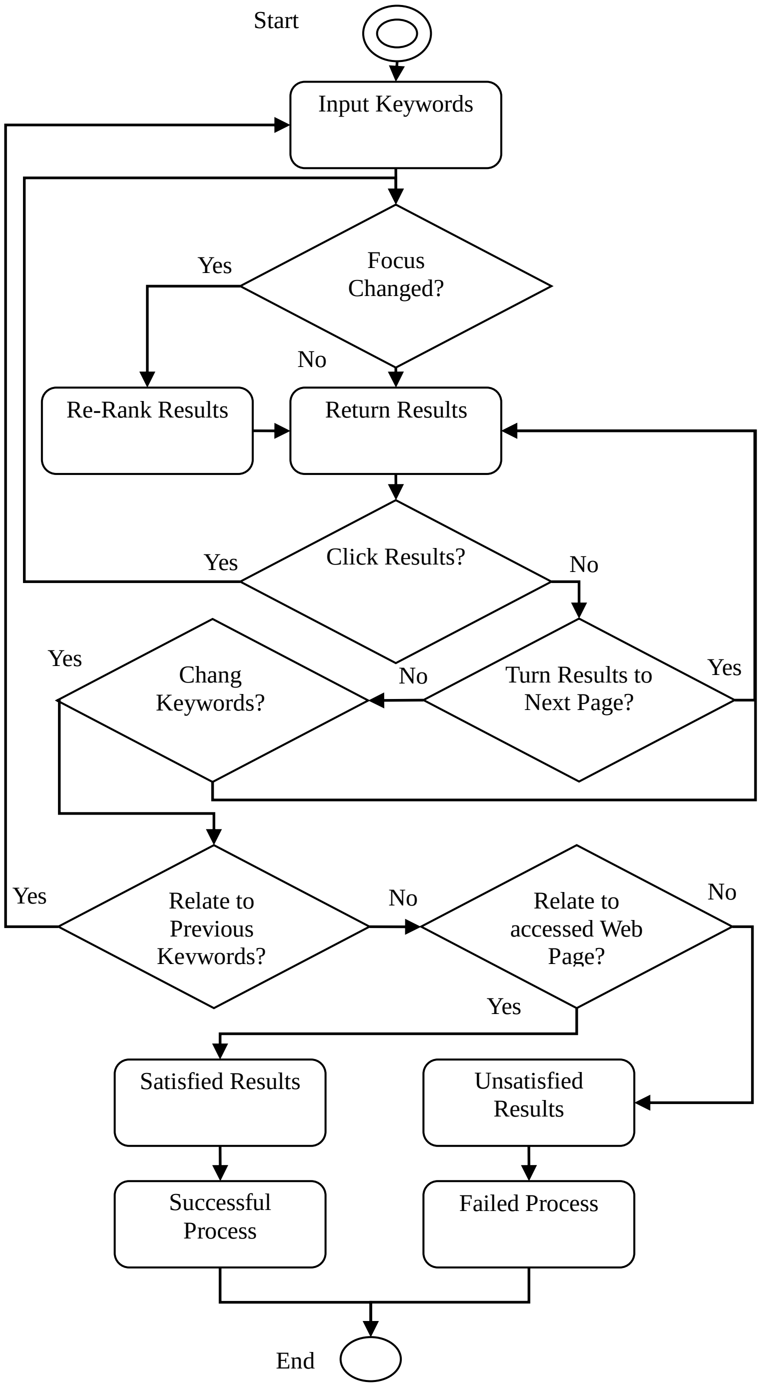

Figure 4-8. Flowchart of optimized process recommendation

his/her focus. After checking the results, the user may have three options: clicking results, turning results to next page, or changing the keywords. If the option of the user is turning results to next page, it means that the seeking process is ongoing, and the system will return the other results page to the user. If the option is clicking results, the system will analyze the clicked web page, and rearrange the top tag group of the user, in order to check if the user's interest focus changed. If the option of the user is changing keywords, the system will check the new keywords in order to confirm if the new keywords relate to the previous keywords based on WordNet. If the new keywords are the synonym of the previous keywords, it also means that the current seeking process is ongoing. In this case, the system will return the new results to the user. But if not so, the system will regard the current seeking process to end, and a new seeking process is to begin. And then, the system will check the relation between the last inputted keywords and the last accessed page. According to WordNet, if the keywords and page are marched based on criteria, the last accessed page will be regarded as a satisfying result for the seeking process. In reverse, this seeking process will regarded as an unsuccessful process by the system. For a successful seeking process, it will be used for optimizing seeking process.

Table 4-1. Search results comparing of search engines

		WordNet			other
		island	coffee	language	
Google	P1	0	0	9	1
	P2	0	0	9	1
	P3	0	0	10	0
Yahoo Japan	P1	0	0	9	1
	P2	0	0	9	1
	P3	0	0	8	2
Bing	P1	0	0	8	2
	P2	0	0	5	5
	P3	0	0	9	1
Baidu Japan	P1	0	0	9	1
	P2	0	0	9	1
	P3	0	0	10	0

Investigated on Apr. 17, 2012

Some investigations indicated that most of users view first three pages of search results provided by search engines [57] [4], therefore, we tried to compare search results of some search engines (e.g. Google, Yahoo Japan, Bing, Baidu Japan) based on WordNet. Table 4-1 represents the search results about keyword java. Based on WordNet, there are three clusters that relate to the keyword java, which are an island of Indonesia (island), a kind of coffee namely java (coffee) and a programming language of java (language). As shown in Table 1, the P1 to P3 indicate the search results page number; the numbers under the island indicate the number of search results related to an island of Indonesia; the numbers under the coffee indicate the number of search results related to the coffee namely java; the numbers under the language indicate the number of search results related to programming language of java; the numbers under the other indicate the number of search results except the forgoing clusters. The results denote the search results of these search engines almost focus on programming language of java till the third page. Therefore, if a user wants to search the information about java except programming language, the user will hardly retrieve the wanted information. This problem is solved by the proposed system.

For example, when a user inputs keyword java to search something, the proposed system will return there clusters search results according to the Synset relations summed by WordNet. If the user's interest focuses on an island of Indonesia, the system will return the results related an island of Indonesia to the user at the page's top, and the other results will be shown under them. When the user adds or changes the keywords, for instance, adding keyword Jakarta or changing keyword java to Javanese Javan, the system will check the new keywords. If the new keywords relate to the Synset of java in WordNet, which means the current seeking process is not end. In this case, the system will continue the seeking process, and extract reference user group by analyzing seeking log that relates to the Synset of java in WordNet. Then, the system will extract satisfied results from successful seeking process of reference user group. Conversely, the system will regard the current seeking process to end, and a new seeking process is to begin. In this case, the system will detect the transition of the user's focus, and extract new reference user group by analyzing the seeking log that relates to the Synset of the new keywords in WordNet.

In the other hand, the processed system will also check the top tag group that is based on Hatena Bookmark Tag, and detect the transition of the user's focus from the change of top tag group. For example, at the first, there is a tag group related to the java island of Indonesia in a user's top tag group. When the user search something about java, the system will return the results related to the island of Indonesia to the page's top. After some time, if the tag group of the java island is removed from the top tag group, and the other tag group that relates to a kind of coffee namely java enter into the top tag group, the system will also re-rank the search results and put the results related to coffee at the page's top.

In summary, the proposed approach not only considers the relation among users, but also considers the individuality of each user. As described in Section 4, by a collaborative means, the proposed system can extract the successful seeking process from a reference user group, at the same time, the individuality of each user is also considered by the Optimized Process Recommendation Algorithm.

In this algorithm, a top tag group has been designed to describe the individuality of users, in which a number of tags are arranged according to their high posterior probability. The transition of top tag group can reflect the transition of users' interest focus. Therefore, the priority of top tag group is higher than the reference user group. This solution means that the priority of individuality is higher than commonness in our proposal.

4.6.2 Simulation

Figure 4-9 shows a simulation of seeking process. The vertical axis denotes the posterior probability of tags, and the horizontal axis denotes the time line. The upper case “S” and “L” on horizontal axis represent shorter time point and longer time point respectively. The Tags A to E denote that such as these tags are focused by the user. Perceptibly, if a tag has a high posterior probability, the selected probability of the web pages that belong to the tag will be higher than the other web pages.

In our study, we define a top tag group in such a way that if the posterior probability of a tag is in the top three, the tag is regarded as a member of top tag group. At a time point, if the posterior probability sequence of members changed, this time point is regarded as a shorter term point. If the member of top tag group changed, this time point is regarded as a longer term point. A term between two adjacent shorter

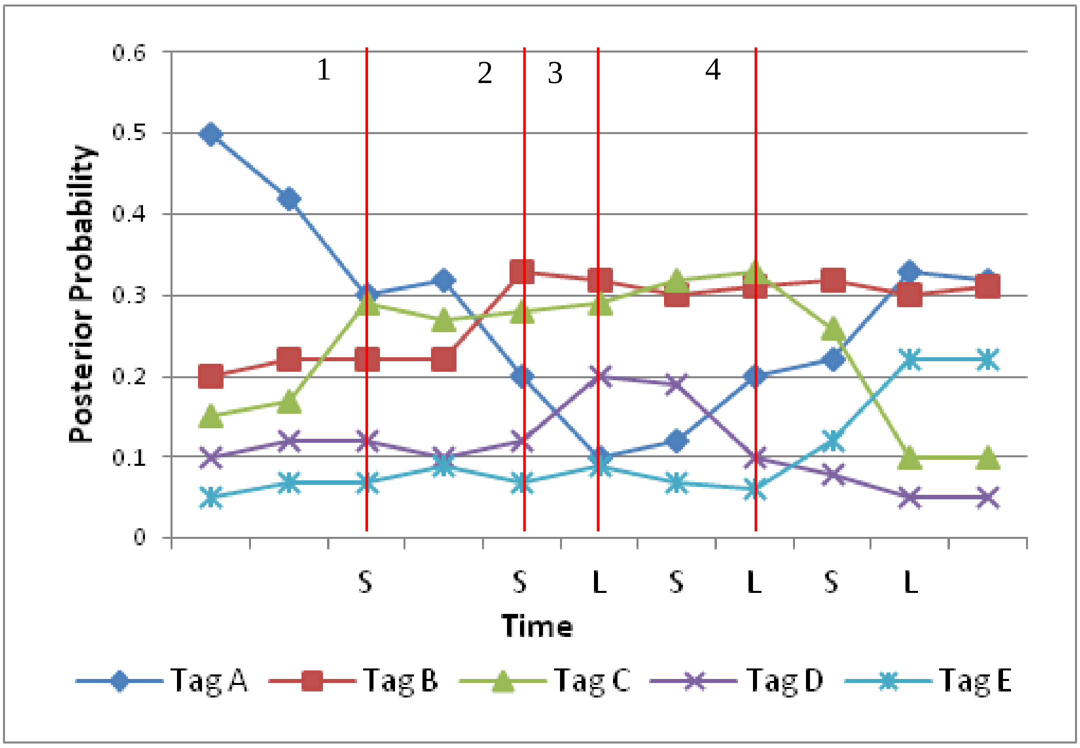

Figure 4-9. Simulation of seeking process

term points is regarded as a shorter term. Similarly, a term between two adjacent longer term points is regarded as a longer term. As shown in Fig 8., the period from Line 1 to Line 2 is a shorter term, and the period from Line 3 to Line 4 is a longer term.

If there is no change of top tag group, it means the target user's interest focus has no change. If the sequence of top tag group changed, it means the target user's interest focus has somewhat change. If the member of top tag group changed, it means the target user's interest focus has a big change. Therefore, detecting such as these changes will benefit to regulate the recommendation results, and help to improve the precision of recommendation.

4.7 Summary

In this chapter, we have introduced an optimized process recommendation model for optimizing information seeking process. In our proposal, we defined a set of measures to describe the patterns of users' information access behaviors, and then an optimized process recommendation algorithm was used to extract and optimize the information seeking process for target users from their reference user group based on the similarity of behavior patterns.

Different with the traditional recommender systems that provide the final results to users directly, our proposal focused on recommending a seeking process to users, because the seeking process can help users to apperceive the knowledge contained in the recommended results gradually. Furthermore, the individuality and commonness are integrated in our approach, which could make the recommendation results more suitable to target users. Finally, based on our previously developed gradual adaptation model, this approach gradually adapted to a target user's needs and information access behaviors so that it benefits to improve the precision of recommendation.

5 Recommendation for Individual Learning Support

As an application of Gradual Adaptation model and Optimizing Process Recommendation model, a prototype system namely Goal-Driven process navigation for individualized learning activities is implemented. In this chapter, after introducing the concept of Goal-Driven Process Navigation and system architecture, .the experimental evaluation will be discussed at the end.

Two Chinese idioms can be used to describe the relation of learning process and result. One is "getting twice the result with half the effort," and another is "getting half the result with twice the effort." A lot of studies indicated that different learning process led to different results in students' learning activities [7] [8] [58]. It is assumed that if the learning process of students is applicable to be regulated and navigated with a suitable principle, more efficient and better results can be expected.

In order to help students to improve their learning efficiency, we propose an integrated approach to optimize learning process. In this study, students are assumed to be divided into two groups: one is called reference student group, in which students have the successful experience in previous learning processes, and another is called target student group, who are the lower-performing students than the reference student group. With our proposed approach, we extract learning patterns of both reference student group and target student group based on their log data, and infer the target students' learning goals by analyzing their current learning actions that are defined as an operating unit in a learning activity, and comparing their learning processes with the reference students whose learning patterns are similar to the target students. Moreover, according to the analysis results of reference student group, those learning actions which may be more suitable to the target student will be chosen as his/her next learning action in the optimized learning process navigations to accomplish a specific learning goal.

Based on these, the working flow can be described as follows. At the first step, students' learning patterns are extracted from the access log data using the clustering method. Then, the reference student group can be built for a target student by comparing the similar learning patterns at the second step. At the third step, a Bayesian network of learning actions is created from the reference student group according to the posterior probability of learning actions, which could be viewed as a set of choices for the target student as his/her next learning action during the learning process navigation. In addition, the selection of the target student from the learning action choice set is regarded as a feedback, which is used for the Gradual Adaptation Model (GA model) proposed in our previous study [1] to improve the recommendation results. In this study, the system architecture for goal-driven process navigation is further developed based on a ubiquitous learning environment empowered by Internet of Things (IoT), so that context data, such as location, situation, can be collected and analyzed, which consequently enhance the GA model and the proposed system as well. Our proposed approach provides a goal-driven navigation of optimized learning process to students, and this solution does not only consider the relation of learning contents, but also take into account of the individual difference of students [[59]]. Therefore, it can be expected to help students to know what need to learn, and furthermore, let them apperceive how to learn.

5.1 Learning Activity and Activity Course Model

In this section, we introduce some definitions of our study firstly, ant then describe the architecture of activity course [60].

5.1.1 Basic Definitions

In this sub section, we introduce the definitions of learning operation, learning action, learning activity, learning process, learning pattern and their relations.

Definition 5.1 (Learning Operation) A learning operation is the smallest operating unit in learning. For example, reciting (new foreign language words) or writing (new foreign language words) is regarded as a learning operation respectively.

Definition 5.2 (Learning Action) A learning action is defined as an operating unit of learning activity. A learning action consists of a set of learning operations. For example, learning new foreign language words is regarded as a learning action, and the above two learning operations of reciting and writing can be used for the learning action of learning new foreign language words.

Definition 5.3 (Learning Activity) A learning activity is defined as an educational process or procedure intended to stimulate learning through actual experience, which consists of a series of learning actions that constitute a purposeful learning process with a certain sequence and time span [Chen, 2010]. For example, in a foreign language lesson, a learning activity consists of the learning actions - learning new words, learning new programs, learning text, doing exercises and doing quiz. But the sequence of learning actions can be optimized so that it can suit to target students.

Definition 5.4 (Learning Process) A learning process can be regarded as a learning activity or a series of learning activities with a specific sequence. On the other hand, the learning activities can be regarded as a purposeful learning process. Further, the learning process is a series of actual and successive learning behaviors based on the corresponding learning activities.

Definition 5.5 (Learning Pattern) The learning pattern is a composite of traits or features characteristic of an individual or a group in learning process. In our study, such as the learning time, learning sequence, learning situation are used to describe the learning patterns.

5.1.2 Learning Activity and Activity Course

An activity course is based on corresponding learning activities. Therefore, an activity course is designed for achieving a series of learning motives, which is made up of a

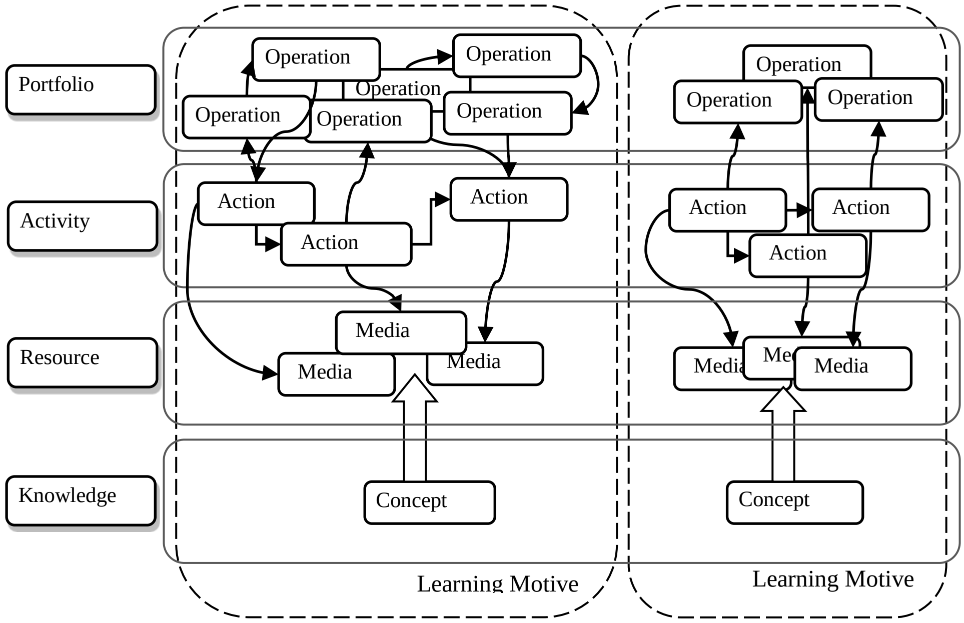

Figure 5-1. The architecture of the activity course

chain of learning activities (lessons). In order to achieve a learning motive, a series of learning actions are arranged in sequence according to the learning principle. Figure 5-1 shows the architecture of Activity Course. It consists of Knowledge, Resource, Activity and Portfolio. Students need to use media and complete operation according to the requirement of learning actions and their sequence. To learn a concept is regarded as a goal, which can be completed by a series of different learning actions.

The course is the component of the curriculum, and can be regarded as a sequence of the students' needs and experiences. As a curriculum, it could be divided into four layers or facets, each of which shows a specific facet of a curriculum. These four layers of a course are described as follows:

- **The knowledge layer** contains the learning concepts of the course, which comes from the domain knowledge. The sequence of the knowledge map can give the direction of connecting the resource to support the course implementation.

- **The resource layer** represents the resource to support the transmission of the knowledge, which is the resource base for a teacher to generate a course, namely the media of the knowledge.
- **The activity layer** is the core part for the practice of the learning and teaching activity, which is always designed by the teacher and followed by the students. In this thesis, we focus on learning activity, but not on teaching activity. In order to help students to accomplish the learning of concept, a series of learning actions are designed by teacher with a specific sequence. Each learning action relates to a special learning resource and learning operation. The activity is mostly determined with one's different value judgment.
- **The portfolio layer** consists of the student's outcome such as a report or discussion record in a forum with timestamp from the learning operation. It also includes the assessment and even the material generated in the learning activity. These materials are traditionally used to give the assessment of the student.

In order to achieve a learning goal, a student can conduct the learning activities that are provided by an activity course, and the four layers can be used to support activity course so as to help students improve learning effectiveness.

5.1.3 Activity Course Model

The proposed activity course model is based on the well-known activity theory. As described in the above, a learning activity can be divided into a series of learning actions, and a learning action can be further divided into a series of operations. All of the activities, actions and operations are organized with a specific sequence. In a normal LAMS (Learning Activity Management System), a learning activity is designed for realizing a learning motive. As shown in Figure 5-2, we give a

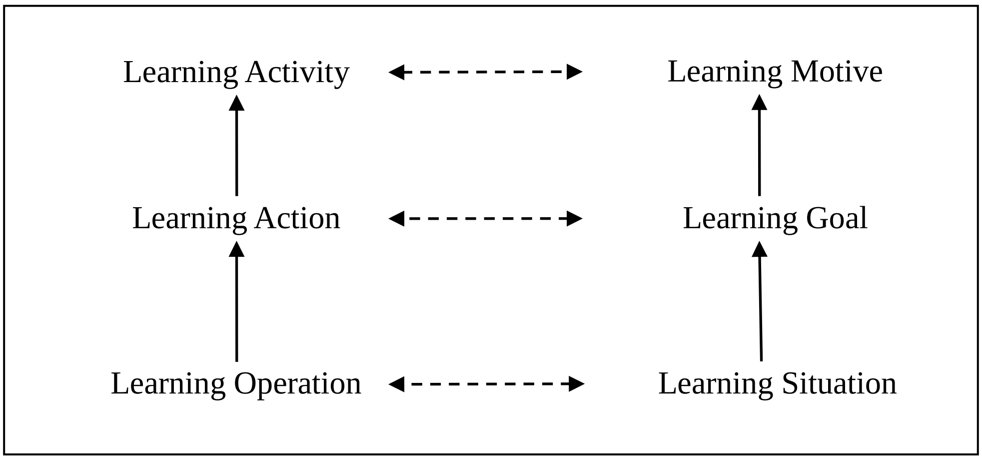

Figure 5-2. Conceptual view of the activity course model

conceptual view for activity course model. In this figure, a learning activity is used to achieve a corresponding learning motive, a learning action is used to achieve a corresponding learning goal, and a learning operation is done based on the learning situation of the student. Learning situation is regarded as an indicator to describe learning operation. When a student is in a learning situation, he/she can do the corresponding learning operation.

In this model, the learning motive is used to describe what a learner wants to learn, and it may correspond to one or plural learning goals. A learning activity is paired with a learning motive. It consists of a series of learning actions that a leaner may engage in. A learning action contains a series of specific practical operations. It guides the learner what he or she needs to do at what time. For example, it can remind an English learner to memorize new words in early morning, to learn English grammar in the morning, and to do exercise in the afternoon. Furthermore, for the learning action of memorizing, transcribing and reciting new words are the detailed learning operations in the learning action.

As described in the above, learning activities are regarded as a learning process with a specific purpose and sequence. Learning actions belong to a corresponding learning activity. Both a learning activity and the related learning actions can be utilized to extract information on user contexts to create the user model (specifically the user profile and group profile if available). They are recorded based on the activity course model as a kind of metadata, which can be used to detect learners' needs, and to extract successful experience as well. The learning activity contains a sequence of learning actions, and the sequence includes metadata, such as time, actors and contents, etc. For example, from viewing lecture video and uploading homework to reviewing the log is a time sequence in a course. Finally, the learning content is paired with the learning operation. On the other hand, by comparing the ongoing learning process with the reference group, we can infer the learning goal of a target student. Therefore, this mechanism can be used to detect students' learning goal, and navigate the next learning operation to target students.

5.2 Concept of Goal-Driven Process Navigation

In this section, after describing basic concepts about the learning process, we introduce the concept of goal-driven process navigation for individualized learning activities.

5.2.1 Purposeful Learning Process

Purposeful learning is activity-based in terms of students applying what they learn through completing assignments or specific tasks related to the assignment [61]. In this study, a learning action consists of a series of operations, and a learning activity consists of a series of learning actions with a specific sequence. As described in the above, a learning activity can be regarded as a purposeful learning process. A standard

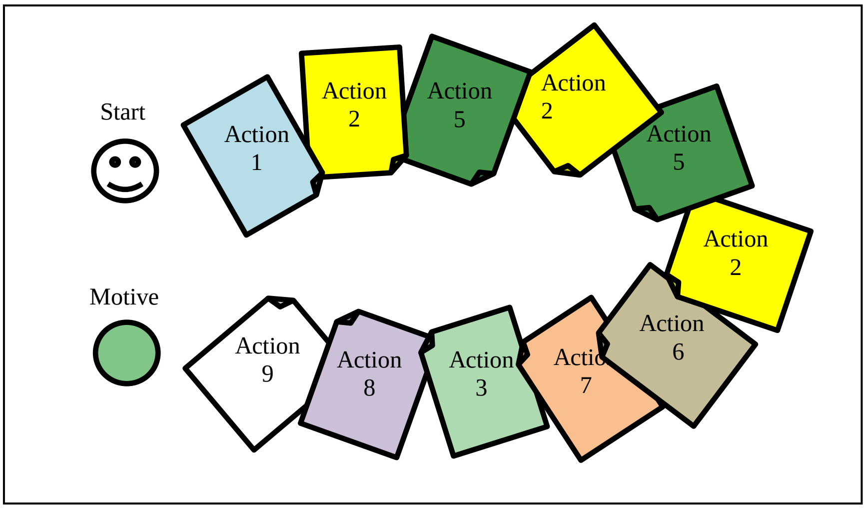

Figure 5-3. An example of learning action process

learning process is given out in a learning activity for a learning goal, and it can be used by students with a free sequence. Figure 5-3 shows an example of how a student takes actions for a learning activity. In this example, the standard learning process consists of Learning Action 1, Learning Action 2, …, Learning Action 9 in a certain sequence. A student can do it with a customized sequence. For example, Learning Action 2 is taken three times, and Learning Action 5 is taken two times.

In order to assess a learning process, we can design a quiz at the end of each learning activity. The performance of the quiz is used to describe the effect of the learning process on a student. In a whole learning process, each learning action may have different contribution to the performance of the quiz, and the access times of learning actions are regarded as a weight for calculating their contribution. The detail of contribution will be discussed in the next section.

5.2.2 Goal-Driven Process Navigation

For a learning process, a different input can lead to a different output. In order to help

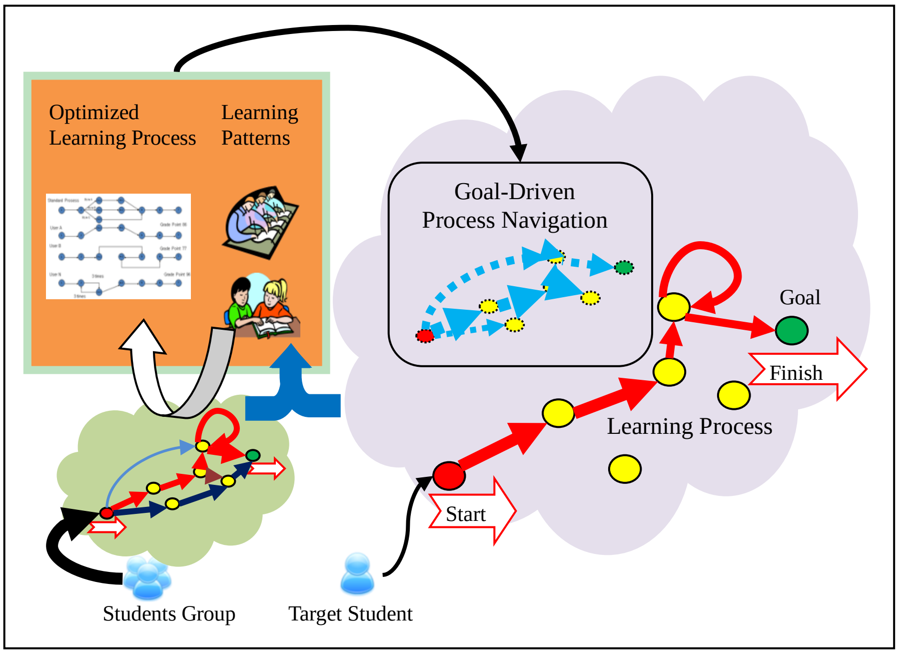

Figure 5-4. Illustration of goal-driven process navigation

students to obtain their expectant output, we need to regulate the learning process, and make it to fit to a target student. The goal-driven learning process optimization approach is used to solve the problem.

Rosenblueth et al. thought that active behavior may be subdivided into purposeless and purposeful classes, and the purposeful behavior means that it can direct to the attainment of a goal while the purposeless behavior cannot direct to the goal [62]. In this study, by analyzing learning actions of students, we can infer their purpose by mining the log data of other students who have similar learning patterns. And then, we can recommend a set of potential next learning actions to them. After the student selected one of the recommended learning actions, this selection can be further used to predict next learning behavior. Hence, we call our approach as goal-driven learning process navigation.

As shown in Figure 5-4, learning patterns of students are extracted by clustering at first. Based on the similarity of learning patterns and the level of experience, a reference student group is extracted for a target student. After the target student started a learning process, his/her current learning action is used to detect his/her learning goal. Based on the collaborative filtering algorithm, a Bayesian network is built. Using the learning process log data of reference students group, the posterior probabilities of next learning actions are calculated, and then a set of choices for next learning actions are delivered to the target student by descending order of posterior probability. After the target student selects the next learning action, new choices will be delivered to him/her till the target student accomplishes the learning process.

5.3 System Architecture

In this section, the system architecture, its major functional modules, and an integrated algorithm are introduced and described.

5.3.1 Overview

As shown in Figure 5-5, the system of goal-driven process navigation consists of User Interface, Situation/Context Analyzer (SCA), User Profile Creator (UPC), Learning Pattern Analyzer (LPA), Learning Process Optimizer (LPO) and Gradual Adaptation Recommender (GAR), in addition to Learning Activity Management System (LAMS) and Search Engine (SE). A specialized User Interface is designed for goal-driven process navigation, which can be used to receive the access behaviors data including location data such as GPS information of students. The SCA is used to analyze the location data transmitted by User Interface, and save the extracted situation/context metadata to the database of Access Logs. The UPC is used to create user profiles that are used to analyze learning patterns. Students can access the goal-driven process navigation, and Learning Activity Management System (LAMS) and Search Engine

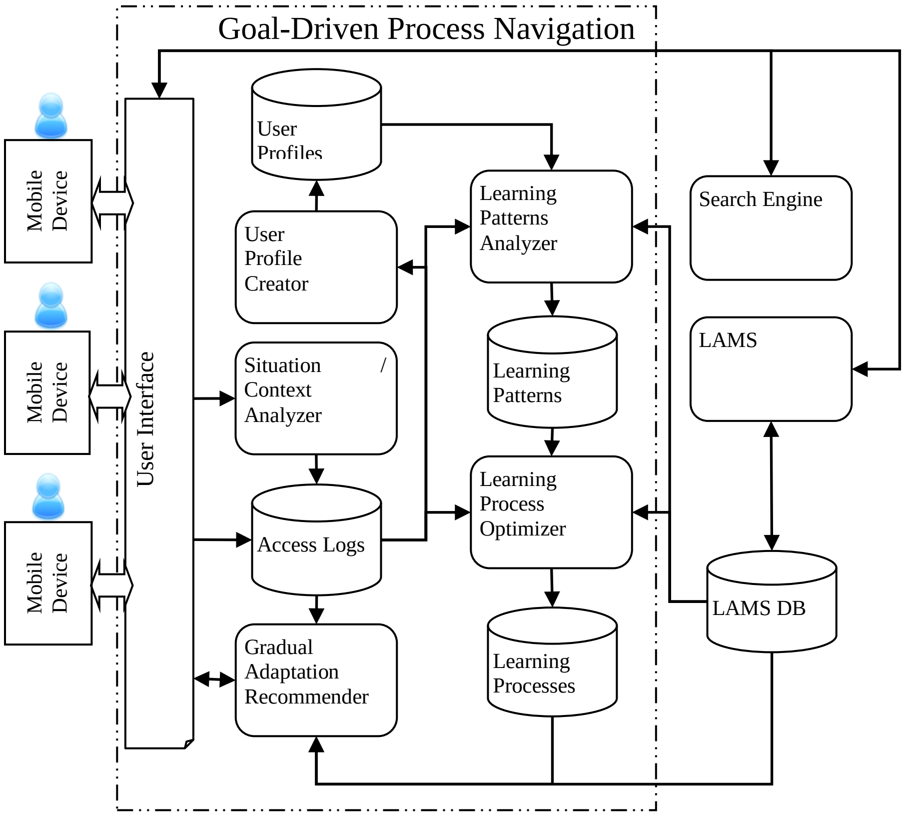

Figure 5-5. System architecture of goal-driven process navigation

(SE) as well through the User Interface. All of access logs are recorded into the database of Access Logs.

Moreover, the LPA is used to extract learning patterns of students. By analyzing access logs and the standard learning process of LAMS, the extracted learning patterns are recorded into the database of Learning Patterns. The LPO is used to optimize the learning process from the reference student group for a target student. According to the similarity of learning patterns, a series of learning actions that have high contribution for the learning performance are extracted from the reference group. Furthermore, the GAR, as one of the core modules in this system, is used to re-rank the recommended learning actions. According to the selection of the target student, it adapts to the learning transition of the target student gradually, and makes the recommended learning process more suitable for the target student. The details of these three modules are introduced in the next sub-sections.

5.3.2 Learning Patterns Analyzer

The algorithm of Learning Patterns Analyzer is shown in エラー! 参照元が見つかりません。. Its detail is described after this figure.

LearningPatternClustering()

Input: student profile $S = \{s_1, s_2, \ldots, s_n\}$
Output: centroids g[k]

1. Initialize cluster
   For data set S, we need to extract k clusters, set initial centroid g = $\{\overline{g_1}, \overline{g_2}, \ldots, \overline{g_k}\}$ from S randomly, and set $G = \{G_1, G_2, \ldots, G_k\}$ from S randomly, where $k < n$;

2. Clustering
   2.1. find the new centroids of each cluster $G_1, G_2, \ldots, G_k$
   for (i = 1; i<k; i ++) {
   for (j = 1; j < number of data in G_i; j ++) {
   $Dis_j = \sum_{i=1}^{k} \sum_{s_j \in G_j} || s_j - \overline{g_i} ||^2$ $\mathrm{Dis_j} = \sum_{i=1}^{k} \sum_{s_j \in G_i} ||s_j - \overline{g_i}||^2$, where $\overline{g_i}$ is the current centroid in G_i.
   }
   if (Dis_j is argmin) {
   set s_j into g[k], where $s_j \in G_i$, s_j is new centroid of G_j
   }
   }
   2.2. reset data set to the nearest cluster
   for (j = 1; j < n; j ++) {
   for (i = 1; i<k; i ++) {
   $Dis_j = \sum_{i=1}^{k} \sum_{s_j \in G_i} || s_j - \overline{g_i} ||^2$, where $\overline{g_i}$ is the current centroid in G_i.
   }
   if (Dis_j is the nearest to new centroid in G_i) then set s_j to G_i
   }
   2.3. Do 2.1 and 2.2 till no data can be reset.
3. Return g[k]

Figure 5-6. Algorithm for pattern clustering

We extract a data set from the learning activity log data to describe students in a learning activity, that is, $L = \{a_i, d_i, w_i, t_i, p_i\}$, where a_i denotes a learning action ID; d_i denotes the distance between the standard learning action and the actual learning actions of students; w_i denotes the access day of a week; t_i denotes the access time; p_i denotes the access situation. We can obtain a_i from the action ID, obtain d_i by comparing the standard learning action and the actual learning action, obtain w_i, t_i from the action time, and obtain p_i from the action service. The detail will be discussed in Section 5.3.5. Moreover, we assume that there exist k learning patterns among n students $S = \{s_1, s_2, \ldots, s_n\}$, and there exist m learning actions for a student in a learning activity. Using K-means clustering, we can obtain k clusters. The pattern tendency of a student can be estimated by the distribution of his/her learning actions in the clusters.

In this algorithm, we divide all of data sets into k clusters at the first step. The second step is calculating new centroids in each cluster. The third step is resetting data into the nearest cluster according to the shortest distance that is from data to the new centroids. This process will be repeated till no data can be reset. The proposed algorithm is based on K-means. By this algorithm, we can obtain the clusters of learning patterns.

5.3.3 Learning Process Optimizer

Based on the results calculated by the LPA, we can obtain some students groups that belong to the same cluster with the target student. It means that a target student is possible to belong to more than one group. By moving out the students who do not have high performance, the reference groups are created for a target student.

In this study, we consider the contribution of learning actions for the learning performance. The basic idea is that a more frequently used learning action is considered to have more effect to students. Therefore, the access number of a learning action is used as a parameter to estimate the weight of a learning action. We use

Equations (10) and (11) to describe the contribution of a learning action.

Equation (5.1) denotes the contribution of learning action $lact_j$ to student s_i. Here, GP_i is the percentage of grade point that the student s_i obtained in a learning activity divided by full marks, $\|lact_j\|$ is the times of action $lact_j$ used by student s_i and $\sum\|lact_j\|$ is the total times of learning actions taken by student s_i in a learning activity.

$$Contribution(lact_j \rightarrow s_i) = \frac{\| lact_j \|}{\sum_{j=1} \| lact_j \|} \times GP_i \tag{5.1}$$

For example, there is a learning action r in a learning activity. A student s_i accessed the learning action r 3 times, and accessed the other learning actions in the same learning activity 7 times. The grade point of the student is 7 points and the full marks are 10 points. The contribution of learning r to the student can be calculated as follow:

$$Contribution(r \rightarrow s_i) = \frac{3}{7} \times \frac{7}{10} = 30\% \tag{5.2}$$

Equation (5.3) denotes the contribution of learning action $lact_j$ to all students. Here, the numerator of the equation denotes the total contributions of action $lact_j$ taken by all students and the denominator denotes the total contributions of all actions taken by all students.

$$Contribution(lact_j) = \frac{\sum_{i=1} Contribution(lact_j \rightarrow s_i)}{\sum_{i=1} \sum_{j=1} Contribution(lact_j \rightarrow s_i)} \tag{5.3}$$

Using these two equations, the contribution of learning actions can be calculated. And then, the learning action which has higher contribution is extracted. Of course, there is a possibility that the access number of a learning action has deviation. We expect using the average access number can avoid this problem. It means that when we calculate the value of contribution, if the access number is bigger than the average access number, this reference student's data will be moved out.

5.3.4 Gradual Adaptation Recommender

The GAR is used to detect and adapt students' learning transition gradually, and then generate the learning process navigation for students. A Bayesian network is created in the GAR.

As shown in Figure 5-7. An example of learning process navigation, the dotted lines denote a standard learning process, and the solid lines denote an actual learning process used by students. Thickness of the line denotes the utilization rate. The thinner line means that its application rate is smaller than the thicker line. Here, we assume there are k learning patterns $G = \{g_1, g_2, \ldots, g_k\}$ for students. $A = \{a_1, a_2, \ldots, a_n\}$ are the learning actions' access number of the reference group students. In detail, a_1 denotes the access number of Learning Action 1 after the current learning action of a target student. Then, Equation (5.4) is used to calculate the posterior probability.

$$P(G = g \mid A = a) = \frac{P(G = g)P(A = a \mid G = g)}{P(A = a)} \tag{5.4}$$

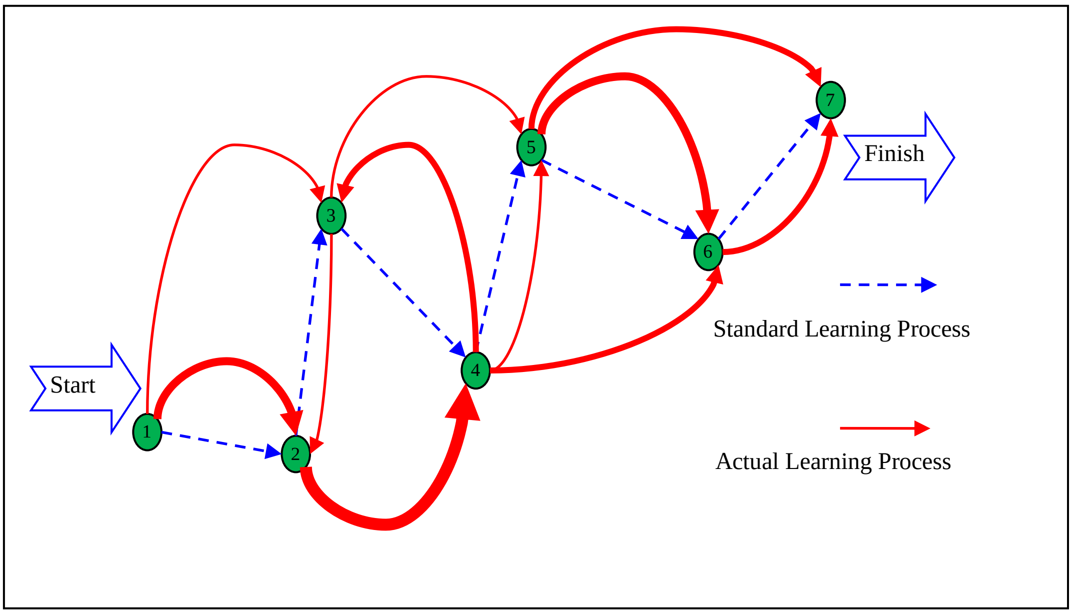

Figure 5-7. An example of learning process navigation

In Equation (3), G denotes the learning pattern type, A denotes the target learning action; $P(G{=}g \mid A{=}a)$ denotes a posterior probability of the target learning action a is selected by the students whose learning pattern is g; $P(G = g)$ denotes a prior probability of a target learning action can be selected by the students whose learning pattern is g; $P(A{=}a \mid G{=}g)$ denotes the probability of learning action a is accessed by students whose learning pattern is g; $P(A{=}a)$ denotes the probability of the learning action a is accessed by all of students. When we want to calculate the posterior probability of learning pattern g_i and learning action a_j is selected, Equation (5.4) can be changed to Equation (5.5).

$$P(G = g_i \mid A = a_j) = \frac{P(G = g_i)P(A = a_j \mid G = g_i)}{P(A = a_j)} \qquad (5.5)$$

Since the prior probability can be obtained from the reference group, Equation (5.5) can be further expressed as Equation (5.6).

$$P(G = g_i \mid A = a_j) = \frac{\frac{1}{\| G \|} \times \frac{a_j}{\sum_{a_j \in g_i} a_j}}{\frac{a_j}{\sum a_j}} \qquad (5.6)$$

In Equation (5.6), $\|G\|$ denotes the number of learning patterns. Based on Equation (5.6), we can *calculate* the posterior probability of a target student who belongs to learning pattern g_i, and selects learning action a_j. After the posterior probability of learning action is calculated, the results are recommended to the target student by the descending order of posterior probability.

For example, there are 10 learning patterns were extracted in an activity course, for a target student s whose preferred learning pattern is g, there exists a reference group for the target student. After the target student accessed a learning action namely current learning action, the proposed system will find the next learning action accessed the *reference* group. We assume the next learning action is a. If the access times of learning action *a* is 30 and access times of the other learning action is 20 in the reference group, moreover, the access times of learning action *a* is 40 and the access times of the other learning actions is 60 in the whole students, the posterior probability of learning action *a* can be calculated as follow:

$$P(G = g \mid A = a) = \frac{\frac{1}{10} \times \frac{30}{50}}{\frac{40}{100}} = 0.15 \qquad (5.7)$$

After the target student selects a learning action, the system is then repeated with the above *process*: calculates the posterior probability of next learning action, and deliveries the results to the target student.

5.3.5 User Profile Based on Learning Activities

In order to help student to achieve a learning goal effectively, we need to design a learning activity for the student. Because of the diversity, the student may achieve the learning goal along the design or progress in his/her own favorite pace. When a student uses the proposed system, the access data will be saved as access logs.

The learning activity is modeled as follows. Each learning activity consists of a series of learning actions with a sequence, and a learning action is composed of six elements (partly optional) described as follows.

- Time: when the access starts and ends
- Actor: who is using the learning system
- Content: which content is accessed
- Service: which service is chosen
- Operation: a concrete realization procedure
- Situation: learning status and location of students

```
<activity>
   <action id=”c”>
      <time />
      <actor />
      <content />
      <service />
      <operation />
      <situation />
   </action>
   <action id =”a” />
      …
   < action id =”x” />
```

Figure 5-8. Data structure of learning activity

The basic data structure is shown in Figure 5-8, and its elements are described as the bullet points. Specifically, the service represents a system component of the LAMS, which is used by students. For example, a BBS search service provided by the LAMS, which can help students to search the history of BBS in the LAMS. The situation includes learning status and location data of students. Here, the learning location is a geographical concept, for example, learning in a classroom or on a train. The learning status indicates the environment characteristics of places, for example, learning in a static environment or in a moving environment. It is conceivable that the multimedia learning content of text, audio and video is suitable for being used in the classroom or at home, and the learning content of audio is more suitable for being used while riding on a bicycle. Student’s access logs are recorded, and used to create user profiles. In these profiles, the data such as which action is accessed by which sequence, when and how long it was taken, can be used to describe the past learning activities of a student.

5.5 Experimental Evaluation

A prototype system for experimental evaluation has been built within the Moodle system, a learning content management system. We designed an activity course that consists of 15 learning activities corresponding to 15 weeks. Every learning activity begins on Monday, and ends on Sunday. A quiz is prepared at the end of each standard learning activity. Except review learning activities, there are 8 standard learning activities. Learning process logs of more than 30 students' were used to infer the successful learning activity, in which the grade point is set to be higher than 8 points in a quiz (full is 10). Figure 5-9 shows a learning activity course, in which the Step 1 to Step 7 denote a standard learning process, the last learning action of user is shown under the standard learning process, the recommended learning actions and their posterior probability are shown in a table under the last learning action of user. In this

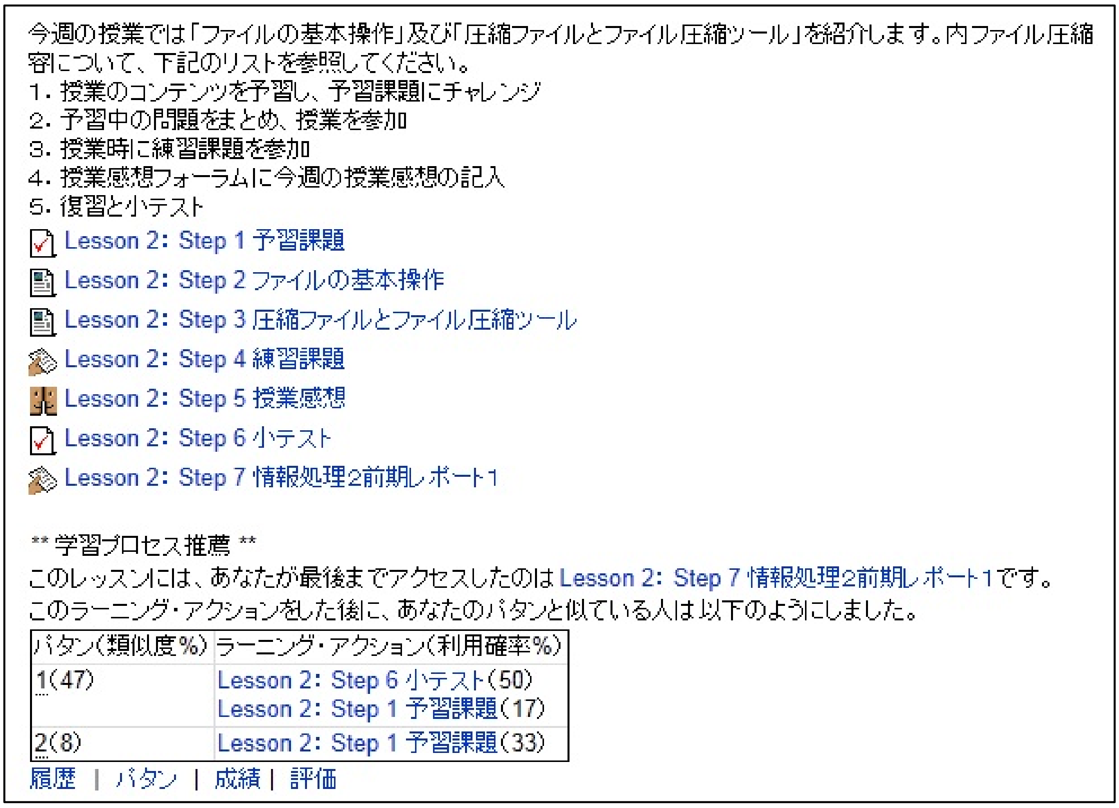

Figure 5-9. The Implementation of Activity Course

table, 1(47) and 2(8) denote the patterns of target user are similar to Pattern 1 and Pattern 2, and the similarities are 47% and 8% respectively. The recommended learning actions are shown at the right of learning patterns. Here, the recommended learning action “Lesson 2: Step 6 (50)” denotes that the utilizing probability of action “Lesson 2: Step 6” is 50% after utilized action “Lesson 2: Step 7” in the users whose pattern is similar to Pattern 1 and the similarity is 47%. Similarly, the recommended action “Lesson 2: Step 1 (17)” denotes that the utilizing probability of action “Lesson 2: Step 1” is 17% in the same reference user group.

By analyzing the log data, three most frequently used learning patterns are extracted, which are described as follows.

- Pattern 1:
  - a) visiting forum/discussion (F/D),
  - b) viewing learning content (C),
  - c) viewing/doing quiz (Q).
- Pattern 2:

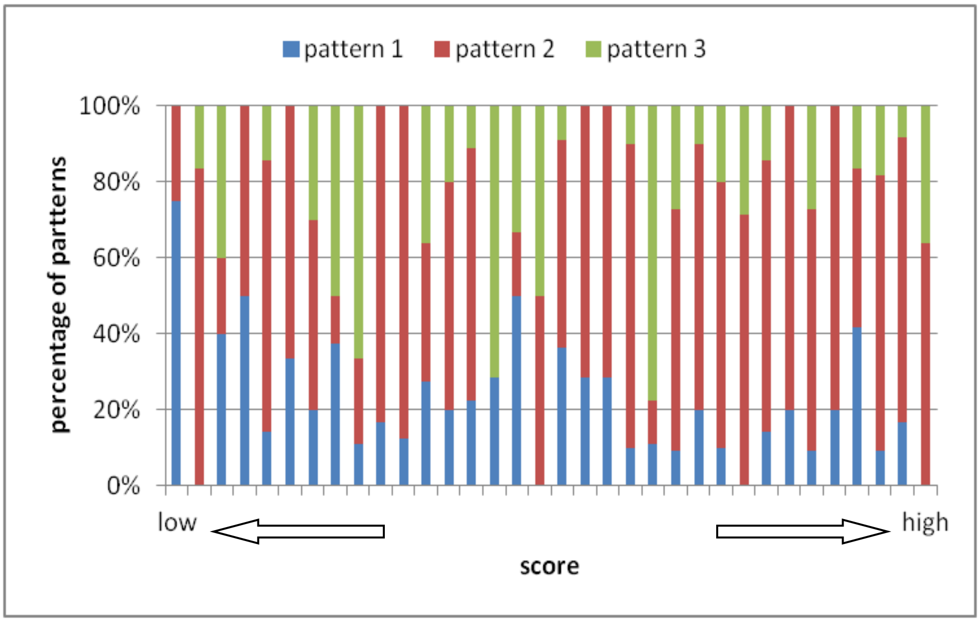

Figure 5-10. Distribution of learning patterns

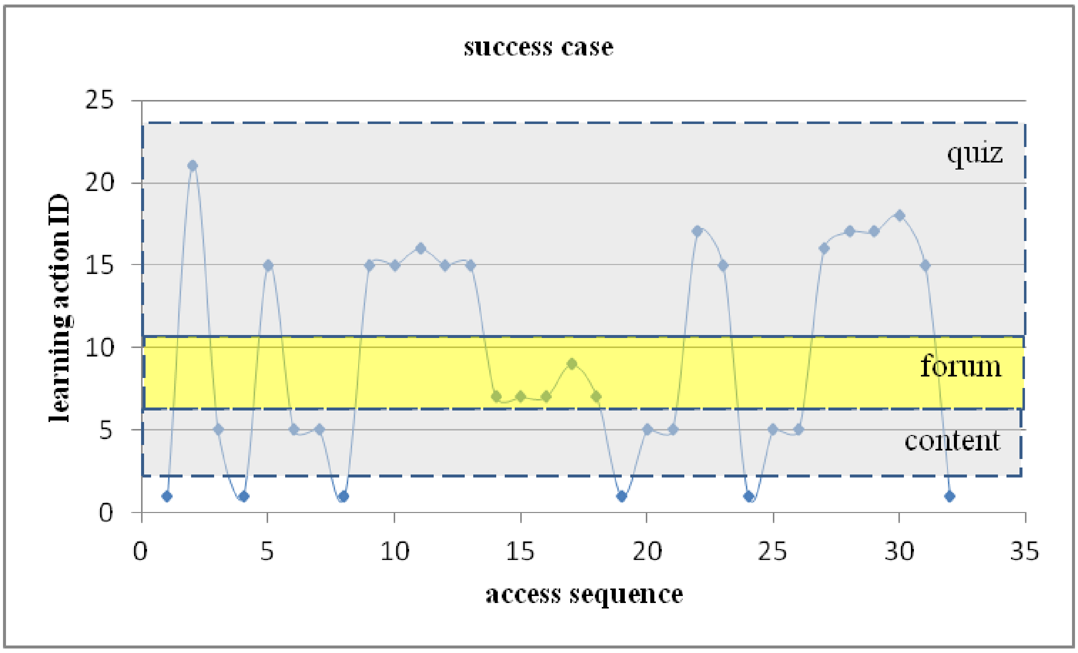

Figure 5-11. Success case

a) viewing learning content (C),

b) viewing/doing quiz (Q),

c) visiting forum/discussion (F/D).

- Pattern 3:

  a) viewing/doing quiz (Q),

  b) viewing learning content (C),

  c) visiting forum/discussion (F/D).

The patterns mentioned above indicate that the actual learning processes of students are different. Figure 5-10 shows the distribution of students' learning patterns. The different color indicates that most students have plural patterns in the whole semester, but most of them have their main patterns or priority patterns. Therefore, we can recommend next learning actions according to the percentage of patterns for learning process navigation. If a student does not select the recommended learning action that is based on his/her learning pattern, our proposed navigation will increase the recommending weight of his/her second learning pattern.

The analysis result proved that the learning achievement can be affected by the

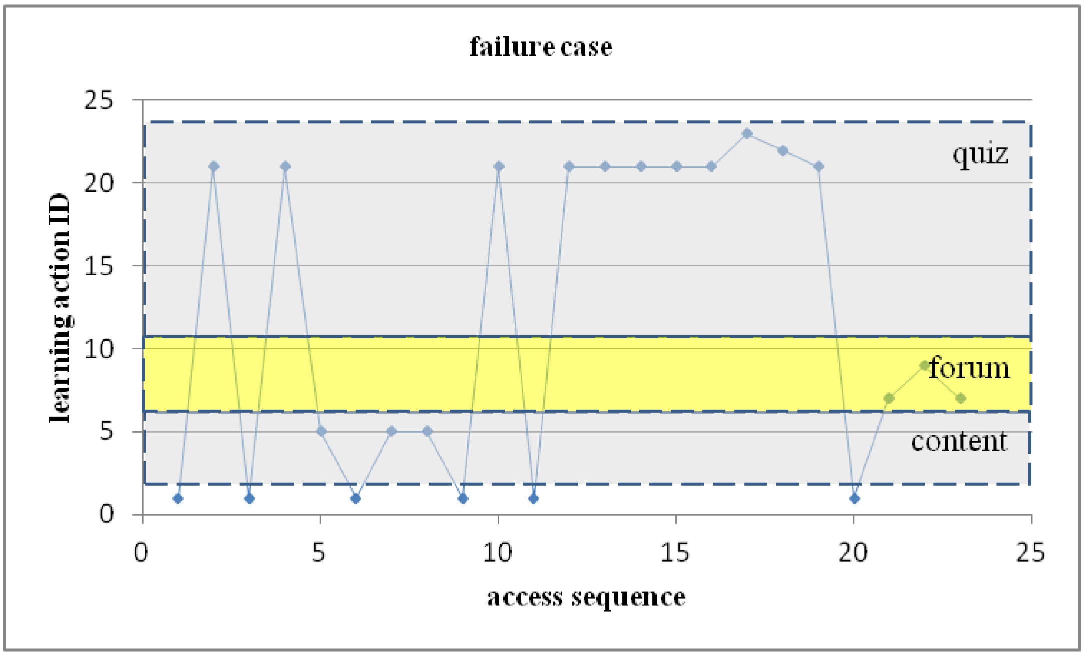

Figure 5-12. Failure case

learning process. In order to describe the relations of learning action and sequence, we set a ID to each learning action. The action ID 1 to 3 denote the other learning actions such as login; the action ID 4 to 5 denote the learning actions of learning detail contents such as text or media resource; the action ID 6 to 13 denote the learning actions about forum and discussion (e. g. viewing a posted article, replying an comment); the action ID 14 to 23 denote the learning actions about quiz. Figure 5-11 shows a success case of learning process, in which high grade point was obtained by the student, where the student started viewing/doing a quiz (Q), and then viewed the learning content (C). After repeating these actions for some times, the student moved to the forum/discussion (F/D), and then returned to do learning actions C and Q. At last, this student obtained a high score at the end of the learning activity. On the other hand, Figure 5-12 shows a failure case of another student's learning process, in which the student obtained a low grade point. This student also started viewing/doing quiz (Q), but almost did not visit forum/discussion (F/D). The score of this student is low

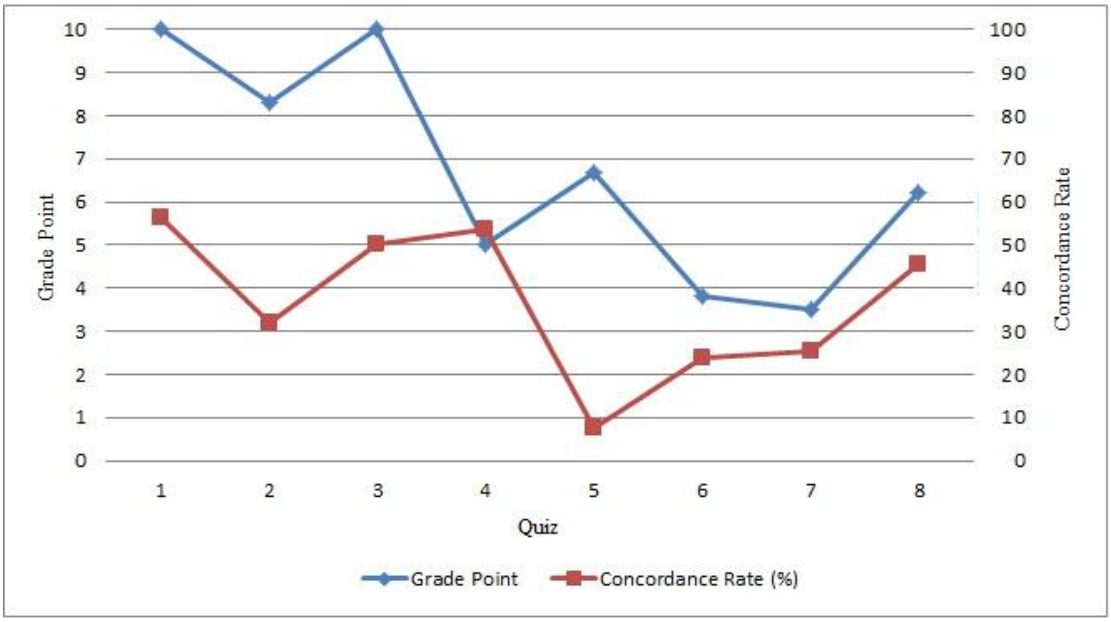

Figure 5-13. The Tendency of Grade Point and Concordance Rate of Process

in this learning activity. Comparing these two students, the results indicate that using forum and attending discussion in an appropriate time can improve the learning performance.

At the same time, the analysis results denote that there exists a similar tendency between grade point and concordance rate (Figure 5-13). Here, concordance rate represents the rate that the recommended learning actions are utilized by target user. For example, if 10 actions are recommended, and the 6 actions are utilized by target user, the concordance rate is 60%.

Finally, we compared the final achievements of students of two classes that took the course in 2010, in which our proposed system was not used, and 2011, in which our proposed system was adopted. The result is shown in Figure 5-14, where A, B, C, D, and F in the horizontal axis represent the grades (A is the highest, and F denotes failure), and the vertical axis represents the percentage of each grade. In Figure 5-14, the grades of Class 2011 are higher than Class 2010 for A, B, D and F, and lower than for C. The reason for the result is that the grade point of a successful learning activity

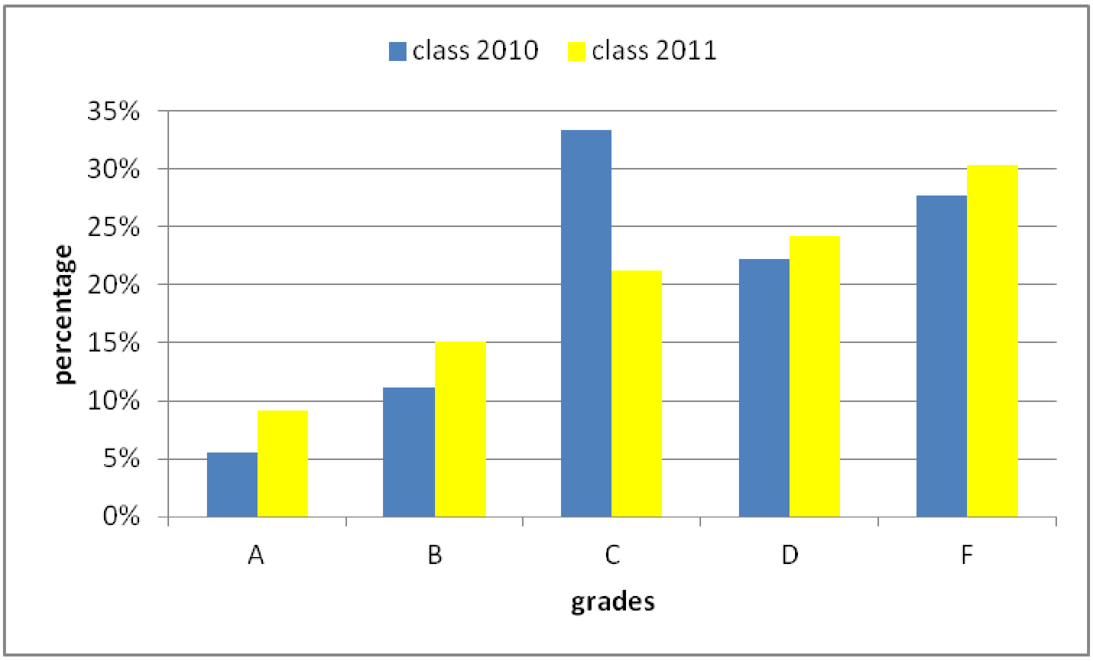

Figure 5-14. Distribution of grades

was set as 8 points, which implies that more target students succeeded in improving their learning performance (more A and B grades) with the help of recommendation and navigation from the prototype system. On the other hand, those students whose usual learning achievements were much lower than 8 seem to be unable to match the recommendation and navigation, which results in more D and F grades. This problem can be expected to be solved by using a variable aimed achievement (not a fixed one, like 8 point in the prototype system).

The results discussed above indicate that our proposed approach and prototype system can be used to improve individualized learning. With the improvement of the proposed system and algorithm, more detailed and precise learning patterns are expected to be extracted, so that more successful and satisfactory learning experience can be shared with each other.

5.6 Summary

In this chapter, we have described an integrated adaptive framework for individualized goal-driven learning process recommendation and navigation in the ubiquitous learning environments. At first, we have described the concept of learning activity and our vision on goal-driven learning process navigation for individualized learning activities. We have introduced the activity course model based on the well-known activity theory, in which the learning goal of a target student can be inferred by comparing the learning processes of the target student with his/her reference student group of similar learning patterns. To show the effectiveness of our proposed framework and approach, we have described the system architecture and its core functional modules, and data structure. Finally, we have shown the experimental evaluation and analysis result, which was based on the prototype system.

6 Conclusion

6.1 Summary of this Study

6.2 Future Work

In this chapter, the contribution of the practical outcomes is summarized firstly. The challenging issues are then discussed. The future work is the end of this thesis.

6.1 Summary of this Study

The age of information explosion bring a new challenge to us. The value of information is inestimable, but the value of information for everyone is different. How to mine for the valuable information for different people is a tremendous challenge. Newton said: "*if I have been able to see further, it was only because I stood on the shoulders of giants*." That say, advice from others may help us overcome our gaps. Therefore, mining successful experience and extracting suitable action process to corresponding users whose behavior patterns are similar to their reference user group, which is the original idea of User-Centric Recommendation based on Gradual Adaptation model and behavior analysis.

In this study, *Gradual Adaptation* model and *Optimizing Process Recommendation* model are proposed. The gradual adaptation model belongs to content-based filtering. By analyzing the access histories of target users, the gradual adaptation model focuses on the individuality of information access behavior, which adapts to target users' transition of needs and behavior of information access. In this model, short, medium, long periods and remarkable, exceptional categories are used to express the difference of user's interest focus. The results of simulation indicate that this classification can reflect different situation of user in information seeking process. Aiming at the different situation, recommending the corresponding information to target user can improve the efficiency of information seeking process.

Based on collaborative filtering, the optimizing process recommendation model pays attention to extract the relations between user and user, user and information seeking process, action (a unit in information seeking process) and action. By analyzing behavior pattern, the relations of users, such as similarities of users are extracted. By analyzing the past successful information access process, the relation of user and information seeking process, for example, a relation of certain information seeking process suits some users of a specific pattern, is extracted. Of course, by analyzing the past successful information access process, the relation of an action and its next action is also extracted. Based on the described above, the proposed approach can adapt to the transition of target user's needs flexibly. When a changing of user's needs is detected, this approach can infer the situation of target user firstly by the gradual adaptation model. Then, a reference user group is extracted according to the relation of user and user secondly. At the third step, successful experiences are extracted according to the relation of user and information seeking process. At the end, a set of next actions that have high posterior probabilities and belong to the reference user group are extracted according to the relation of action and action. According to the descending order of posterior probabilities, the next actions are recommended to target user. The flexibility of our approach is shown in the application namely goal-driven process navigation for individualized learning activities in ubiquitous networking and IoT environments.

As described above, this application is implemented on the Moodle that is an open source of learning management system. There are two classes took the same course, in which the one class didn't use the proposed system, and the other one class used this system. The analysis results denote that the tendency of grade point is similar to the concordance rate. Here, the concordance rate represents a rate that the recommended learning actions are utilized by target user. In other words, when a user chose the recommended learning process, his/her grade point is higher than didn't choose the recommended learning process. Comparing the grade points of the two classes, the results represent that the average grade point of the class using the proposed system is higher than the another class. These results indicate that the proposed approach is effective.

The major features and contributions of our work can be summarized as follows. Firstly, our proposed framework is based on a dynamic Bayesian network that gradually adapts to a target user's needs and information access behaviors. Secondly, an integrated algorithm based on activity theory for extracting behavior patterns has been adopted in the proposed system. Thirdly, a set of measures and equations have been introduced and defined to make the navigated action process optimized, by using the extracted behavior patterns. Our proposed framework and system can be expected to help users to improve their information seeking performance.

6.2 Future Work

Although the above results are exciting, we also noticed that there exist some deficiencies in our approach. For example, the effect is not satisfactory for all of users. Therefore, it is very important to extract rational behavior patterns by experiment and evaluation. We expect that rational behavior pattern can help us to improve the accuracy of calculating user similarity. At the same time, we will improve the adaptation mechanism and algorithm to make them fit to most levels of users.

We will further compare the prototype system with the other recommender systems in order to find deficiencies of our approach, and then improve the proposed system through solving the found deficiencies.

Acknowledgements

My heartfelt appreciation goes to my supervisor Professor Qun Jin whose comments and advices were of inestimable value for my study starting from my Master Degree course at University of Aizu.

I would also like to express my gratitude to all the students from Networked Information Systems Laboratory, for their kind support, encouragement and advice.

Finally, I would like to thank my family in China who helped me to develop and always support my determination to study, and my wife Honghua and my daughter Ziying, my son Yanqing for their love and patience that made it possible to complete this study.

References

[1] J. Chen, R.Y. Shtykh and Q. Jin: "Gradual Adaption Model for Information Recommendation Based on User Access Behavior," *International Journal on Advances in Intelligent Systems*, Vol.2, No.1, pp. 192-202 (2009).

[2] T. Claburn, Google Index Reaches 1 Trillion URLs, http://www.informationweek.com/blog/229211019 (accessed on Dec. 12, 2011)

[3] Kitsuregawa, IEEE APSCC2007 Keynote, Dec. 13, 2007.

[4] BBC News, Search users 'stop at page three', http://news.bbc.co.uk/2/hi/technology/4900742.stm (accessed on Dec. 12, 2011)

[5] http://recsys.acm.org/ (accessed on Sep. 9, 2008)

[6] Washington Post, Rise of the digital information age, http://www.washingtonpost.com/wp-dyn/content/graphic/2011/02/11/GR2011021100614.html (accessed on Dec. 12, 2011)

[7] H.V. King: "Foreign Language Reading as a Learning Activity," *The Modern Language Journal*, Vol.31, No.8, pp. 519-524 (1947).

[8] C.-M. Chen: “Intelligent web-based learning system with personalized learning path guidance,” *Journal Computers & Education archive*, Vol.51, No.2, pp. 787-814 (2008).

[9] L. Floridi: “Is Semantic Information Meaningful Data,” *Philosophy and Phenomenological Research*, Vol.70, No.2, pp. 351-370 (2005).

[10] S. Boell and D. Cecez-Kecmanovic: “Attributes of information, in Proceedings of the Americas Conference on Information Systems,” Proc. AMCIS 2010 (2010).

[11] G. Adomavicius and A. Tuzhilin: “Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions,” *IEEE Transactions on Knowledge and Data Engineering*, Vol.17, No.6, pp. 734-749 (2005).

[12] G. Adomavicius, R. Sankaranarayanan, S. Sen and A. Tuzhilin: “Incorporating contextual information in recommender systems using a multidimensional approach,” *ACM Trans. Inf. Syst*. Vol.23, No. , pp. 103-145 (Jan. 2005).

[13] J. L. Herlocker, J. A. Konstan, L. G. Terveen and J. T. Riedl: “Evaluating collaborative filtering recommender systems,” *ACM Trans. Inf. Syst*. Vol.22, No.1, pp. 5-53 (Jan. 2004).

[14] M. Balabanovic and Y. Shoham, Fab: “Content-based, Collaborative Recommendation,” *Communications of the ACM*, Vol.40, No.3, pp. 66-72 (1997).

[15] J. Srivastava, R. Cooley, M. Deshpande and P.-N. Tan: “Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data,” *ACM SIGKDD*, Vol.1, No.2, pp. 12–23 (Jan. 2000).

[16] L. V. Wel and L. Royakkers: “Ethical issues in web data mining,” *Ethics and Inf. Technol.* Vol.6, No.2, pp. 129-140 (Jun. 2004).

[17] I. Becerra-Fernandez: “Searching for experts on the Web: A review of contemporary expertise locator systems,” *ACM Trans. Internet Technol.* Vol.6, No.4, pp. 333-355 (Nov. 2006).

[18] B. Poblete and R. Baeza-Yates: “Query-Sets, Using Implicit Feedback and Query Patterns to Organize Web Documents,” Proc. WWW2008, Beijing, China, Apr. 2008, pp. 41-48.

[19] R. Kosala and H. Blockeel: “Web mining research: a survey,” *SIGKDD Explor. Newsl*, Vol.2, No.1, pp. 1-15 (Jun. 2000).

[20] W. Fan, L. Wallace, S. Rich and Z. Zhang: “Tapping into the Power of Text Mining,” *Communications of ACM*, Vol.49, No.9, pp. 77-82 (Sep. 2005).

[21] S. Taherizadeh and S. Taherizadeh: “Integrating Web Content Mining into Web Usage Mining for Finding Patterns and Predicting Users' Behaviors,” *International Journal of Information Science and Management*, Vol.7, No.1, pp. 51-66 (2009).

[22] R. Baeza-Yates and p. Boldi: *Advanced Techniques in Web Intelligence - I*, Vol.211, pp. 113-142, Springer Berlin / Heidelberg (2010).

[23] D. Pierrakos, G. Paliouras, C. Papatheodorou and C. D. Spyropoulos: “Web Usage Mining as a Tool for Personalization: A Survey,” *User Modeling and User-Adapted Interaction*, Vol.13, No.4, pp. 311-372 (Nov. 2003).

[24] G. Stumme, A. Hotho and B. Berendt: “Semantic Web Mining State of the Art and Future Directions,” *Elsevier Web Semantics: Science, Services and Agents on the World Wide Web*, Vol.4, No.2, pp. 124-143 (2006).

[25] M. Bilenko and R. W. White: “Mining the Search Trails of Surfing Crowds: Identifying Relevant Websites From User Activity,” Proc. WWW2008, Beijing, China, Apr. 2008, pp. 51-60.

[26] M. Eirinaki and M. Vazirgiannis: “Web Mining for Web Personalization,” *ACM Transactions on Internet Technology*, Vol.3, No.1, pp. 1–27 (2003).

[27] T-W. Yan, M. Jacobsen, H. Garcia-Molina and U. Dayal: “From User Access Patterns to Dynamic Hypertext Linking,” Proc. WWW1996, Paris, France, May. 1996, pp. 1007-1014.

[28] R. Baraglia and F. Silvestri: “An Online Recommender System for Large Web Sites,” Proc. IEEE/WIC/ACM International Conference on Web Intelligence (WI'04), Beijing, China, Sep. 2004, pp. 199-205.

[29] X. Fang and O. R. Lou Sheng: "LinkSelector, A Web Mining Approach to Hyperlink Selection for Web Portals," *ACM Transactions on Internet Technology*, Vol.4, No.2, pp. 209–237 (May 2004).

[30] E. Pyshkin and A. Kuznetsov: "Approaches for Web Search User Interfaces: How to improve the search quality for various types of information," *In JoC*, Vol.1, No.1, pp.1-8 (2010).

[31] V. Klyuev and V. Oleshchuk: "Semantic retrieval: an approach to representing, searching and summarizing text documents," *In: IJITCC*, Vol.1, No.2, pp.221-234 (2011).

[32] V. Klyuev and A. Yokoyama: "Web Query Expansion: A Strategy Utilising Japanese WordNet," *In JoC*, Vol.1, No.1, pp. 23-28 (2010).

[33] R. Y. Shtykh and Q. Jin: "A human-centric integrated approach to web information search and sharing," *Human-centric Computing and Information Sciences*, Vol.1, No.2, pp. 1-37 (2011).

[34] W.M.P. van der Aalst: "The Application of Petri Nets to Workflow Management," *The Journal of Circuits, Systems and Computers*, Vol.8, No.1, pp.21-66 (1998).

[35] R. Agrawal, D. Gunopulos, and F. Leymann: "Mining process models from workflow logs," Proc. EDT 6th, 1998, pp. 469-483.

[36] J.E. Cook and A.L. Wolf: "Discovering Models of Software Processes from Event-Based Data," *ACM Transactions on Software Engineering and Methodology*, Vol.7, No.3, pp. 215-249 (1998).

[37] W.M.P. van der Aalst: "Business alignment: using process mining as a tool for Delta analysis and conformance testing," *Requirements Engineering*, Springer, Vol.10, No.3, pp. 198-211 (2005).

[38] W.M.P. van der Aalst: "Process Mining Put into Context," *Internet Computing*, IEEE, Vol.16, No.1, pp. 82-86 (2012).

[39] W.M.P. van der Aalst: *Process Mining*: *Discovery, Conformance, and Enhancement of Business Processes,* Springer, 2011.

[40] B. F. van Dongen and W. M. P. van der Aalst: "A Meta Model for Process Mining Data," Proc. Conference on Advanced Information Systems Engineering, Vol.161, 2005, pp. 309-320.

[41] M. Liu and W.M. Reed: "The Relationship Between the Learning Strategies and Learning Styles in a Hypermedia Environment," *Computers in Human Behavior*, Vol.10, No.4, pp. 419-434 (1994).

[42] R.M. Felder and L.K. Silverman: "Learning styles and teaching strategies in engineering education," *Engineering Education*, Vol.78, No.7, pp. 674–681 (1988).

[43] S. Graf, S. R. Viola, Kinshuk and T. Leo: "Representative Characteristics of Felder-Silverman Learning Styles: An Empirical Model," Proc. CELDA 2006, Barcelona, 2006.

[44] S. Graf, S. R. Viola and T. Leo: "In-Depth Analysis of the Felder-Silverman Learning Style Dimensions," *Journal of Research on Technology in Education*, Vol.40, No.1, pp. 79-93 (2007).

[45] A. Halstead and L. Martin: "Learning styles: a tool for selecting students for group work," *International Journal of Electrical Engineering Education*, Vol.39, No.3, pp. 245-252 (2003).

[46] R. Huang and G. Busby, Activist: "Pragmatist, Reflector or Theorist? In Search of Postgraduate Learning Styles in Tourism and Hospitality Education," *Journal of Hospitality, Leisure, Sport and Tourism Education*, Vol.6, No.2, pp. 92-99 (2007).

[47] L-w Zhang and H. Guo: *Introduction to Bayesian Networks (in Chinese)*, Science Press, 2006.

[48] .Y. Shtykh and Q. Jin: "Dynamically Constructing User Profiles with Similarity-based Online Incremental Clustering," Special Issue on Intelligent Techniques for Personalization and Recommendation, *International Journal of Advanced Intelligence Paradigms*, Vol.1, No.4, pp. 377–397 (2009).

[49] J. Patton: "Understanding User Centricity," *IEEE Software*, Vol.24, No.6, pp. 9-11 (Nov. -Dec. 2007).

[50] S. Arbanowski, S. van der Meer, S. Steglich and R. Popescu-Zeletin: "I-centric Communications," *Informatik - Forschung und Entwicklung*, Vol.16, No.4, pp. 225-232 (2001).

[51] S. Hamouda and N. Wanas: "PUT-Tag: personalized user-centric tag recommendation for social bookmarking systems," *Social Network Analysis and Mining*, Vol.1, No.4, pp. 377-385 (2011).

[52] A. Yazidi, O-C. Granmo, B. John Oommen, M. Gerdes and F. Reichert: "A User-Centric Approach for Personalized Service Provisioning in Pervasive Environments," *Wireless Personal Communications*, Vol.61, No.3, pp. 543-566 (2011).

[53] D-Y. Cheng, K-M. Chao, C-C. Lo and C-F. Tsai: "A user centric service-oriented modeling approach," *World Wide Web*, Vol.14, No.4, pp. 431-459 (2011).

[54] B.P. Knijnenburg, M.C. Willemsen, Z. Gantner, H. Soncu and C. Newell: "Explaining the user experience of recommender systems," *User Modeling and User-Adapted Interaction*, Online First™ (Mar. 2012).

[55] K. Kuutti: *Activity theory as a potential framework for human-computer interaction research*, In B. Nardi (Ed.), Context and consciousness: activity theory and human-computer interaction. Cambridge, MA: MIT press. 1996, pp. 17-44..

[56] J. Chen, X. Zhou and Q. Jin: "Recommendation of Optimized Information Seeking Process Based on the Similarity of User Access Behavior Patterns," *Personal and Ubiquitous Computing* (Springer, to appear).

[57] Marvist, 92% of Search Engine Users View First Three Pages of Search Results, http://www.positivearticles.com/Article/92-of-Search-Engine-Users-View-First-Three-Pages-of-Search-Results/48338 (accessed on Dec. 12, 2011)

[58] C.-M. Chen: “Ontology-based concept map for planning a personalised learning path,” *British Journal of Educational Technology*, Vol.40, No.6, 1028-1058 (2009).

[59] J. Chen, Q. Jin and R. Huang: "Goal-Driven Process Navigation for Individualized Learning Activities in Ubiquitous Networking and IoT Environments," Journal of Universal Computer Science, Vol.18, No.9, pp.1132-1151 (May 2012).

[60] J. Chen, H. Man, N.Y. Yen, Q. Jin, and T.K. Shih: “Dynamic Navigation for Personalized Learning Activities Based on Gradual Adaption Recommendation Model,” *Lecture Notes in Computer Science (Springer)*, Proc. ICWL2010, 2010, pp. 31-40.

[61] R. Kenedy, and V. Monty: “Dynamic Purposeful Learning in Information Literacy,” *New Directions for Teaching and Learning*, Vol.114, No.1, pp. 89-99 (2008).

[62] A. Rosenblueth, N. Wiener and J. Bigelow: “Behavior, Purpose and Teleology,” *Philosophy of Science*, Vol.10, No.1, pp. 18–24 (1943).