International Journal of Fuzzy Logic and Intelligent Systems, vol. 12, no. 4, December 2012, pp. 270−276
http:// dx.doi.org/ 10.5391/IJFIS.2012.12.4.270 PISSN 1598−2645 eISSN 2093−744X

A Dynamic Ontology-based Multi-Agent Context-Awareness User Profile Construction Method for Personalized Information Retrieval

Qian Gao[1], **Young Im Cho**[2]

[1]School of information, Shandong Polytechnic University, Jinan, 250353, China
[2]College of Information Technology, University of Suwon, Hwaseong, 445-743, Korea

Abstract

With the increase in amount of data and information available on the web, there have been high demands on personalized information retrieval services to provide context-aware services for the web users. This paper proposes a novel dynamic multi-agent context-awareness user profile construction method based on ontology to incorporate concepts and properties to model the user profile. This method comprehensively considers the frequency and the specific of the concept in one document and its corresponding domain ontology to construct the user profile, based on which, a fuzzy c-means clustering method is adopted to cluster the user's interest domain, and a dynamic update policy is adopted to continuously consider the change of the users' interest. The simulation result shows that along with the gradual perfection of the our user profile, our proposed system is better than traditional semantic based retrieval system in terms of the Recall Ratio and Precision Ratio.

Keywords : Ontology, User Profile, Dynamic, Context-Awareness, Fuzzy C-Means Clustering.

1. Introduction

With the explosive growth of internet and storage devices, online document servers abound with a large number of documents. It is an important issue in information retrieval field to overcome a flood of information, and provide users with personalized information. Currently, one of the tackled issues is to offer right and useful information to the end user, which requires a proper method to record the user context and personalized applications so as to tailor the users' needs according to their interests and track their changes. As a solution to this problem, user preference profile approach has been employed to provide personalized search results to each user [1]. The idea of user profiling can be summarized as "one that renders the user with an experience tailored to his/her current situation" [2].

Ontology has been proven to be an effective means for modeling digital collections and user profiling. Ontology in the form of hierarchies of user interests has been proposed by Trajkova and Gauch [3]. Gauch at al. [4] also proposed a system that adapted information navigation based on a user profile structured as a weighted concept hierarchy. The above studies prove that ontology can be a very useful tool, because they may present an overview of the domain related to a specific area of interests and be used for browsing and query refinement.

In this paper, we propose a dynamic multi-agent context-awareness user profile construction method based on ontology to consider the users' preferences over time and domain so as to provide the end user with right and useful information. First, we use a two-layer model to construct the domain ontology; then we use a combination of ontology-driven context-awareness modeling with a dynamically changeable multi-layer tree structure to track the user's reading behaviors on both fixed categories and dynamic events. Third, we use fuzzy c-means clustering method to cluster the user's interest domain to determine whether the user's documents of interest belong to the existed interest-domain, based on which to update the user profile, so as to continuously capture the changes of the users' interests. Finally, we comprehensively use our proposed retrieval system and commercial retrieval system to get more useful information which can meet users' requirements to maximum.

The rest of this paper is organized as follows: Section 2 gives a short review of the related work and contributions of our paper. Section 3 describes the framework of our system and the realization of the user profile construction agent and the retrieval result process and sorting agent. Section 4 gives the simulation based on our proposed method and we arrive at the conclusion in Section5.

2. Related Works

2.1. Research on Ontology based User Profile

Ontology has been a basis for the construction of a user model in several personalized systems ranging from information delivery systems to intelligent tutoring systems. In the field of ontology design, efforts have been made by several research groups to facilitate the ontology engineering process,

Manuscript received Dec. 4, 2012; revised Dec. 24, 2012; accepted Dec. 24, 2012.

*corresponding author : Young Im Cho (ycho@suwon.ac.kr) employing both manual and semi-automatic methods. A. Maedche et al.[5] proposed a framework with the semi-automatic method which incorporates several information extraction and learning approaches, in order to face the discovery of relevant classes, their organization in a taxonomy and the non-taxonomic relationships between classes. Aroyo et al. [6] provided a method for automatic acquisition of user knowledge through an ontology-based dialog system. An ontology-based dialog agent, called OWL-OLM, interacts with the user to analyze the current state of the users' knowledge according to the needs for a particular course task. Zhang et al. [7] proposed a system for constructing user models automatically by monitoring the users browsing behaviors in each session. The system keeps track of the usage logs by means of Semantic Web Usage Log Preparation Model (SWULPM). The user model consists of personal ontology which is represented through concept graph.

2.2. Research on Context-Awareness fuzzy representations

Contextual conditions are equally difficult to define and grasp in ways that are devoid of uncertainty and imprecision. Even content semantics are far too complex to be formally described in a complete or unambiguous manner [8], and needs to borrow further information from context to get a precise interpretation. Ph. Mylonas et al. [9] complemented the ontology-based perspective with fuzzy notions for the representation of user preferences, user context, content semantics, and relations between concepts. They proposed a method for user profiling and personalized retrieval in context which is found in the principles of fuzzy sets and fuzzy relational algebra. The method takes advantage of the available techniques in that area, which are suitable to deal with problems involving fuzzy magnitudes

2.3. Contributions of Our Paper

We construct our system based on a two-layer domain ontology model and use a combination of ontology-driven context-awareness modeling with a dynamically changeable multi-layer tree structure to track the user's reading behaviors. Furthermore we use fuzzy c-means clustering method to cluster the user's interest domain which helps to update the personalized user profile. Our proposed method can better understand the users' interests so as to achieve higher precision ration and recall ration.

3. Multi-Agent Context-Awareness User Profile Model

3.1. Framework of the Multi-Agent Context-Awareness User Profile based Personalized Information Retrieval System

The MAPPIR (Multi-Agent context-awareness user profile based personalized information retrieval system) can be logically divided into three agents: user profile construction agent, retrieval agent and retrieval result process agent. In this paper, we mainly concentrate on the realization of the first and last agents.

User profile construction agent is responsible for the identification of the user devices. Then it judges the probably practical necessity of the user according to the ontology-based user profile or reasoning mechanism. Finally, it updates the documents profile and user profile according to the user's feedback.

Retrieval result process agent is responsible for the processing of the retrieval result by comprehensively considering the retrieval results from our proposed system and the retrieval results from the commercial retrieval system.

The workflow of the Multi-Agent context-awareness user profile based personalized information retrieval system is shown in Fig.1.

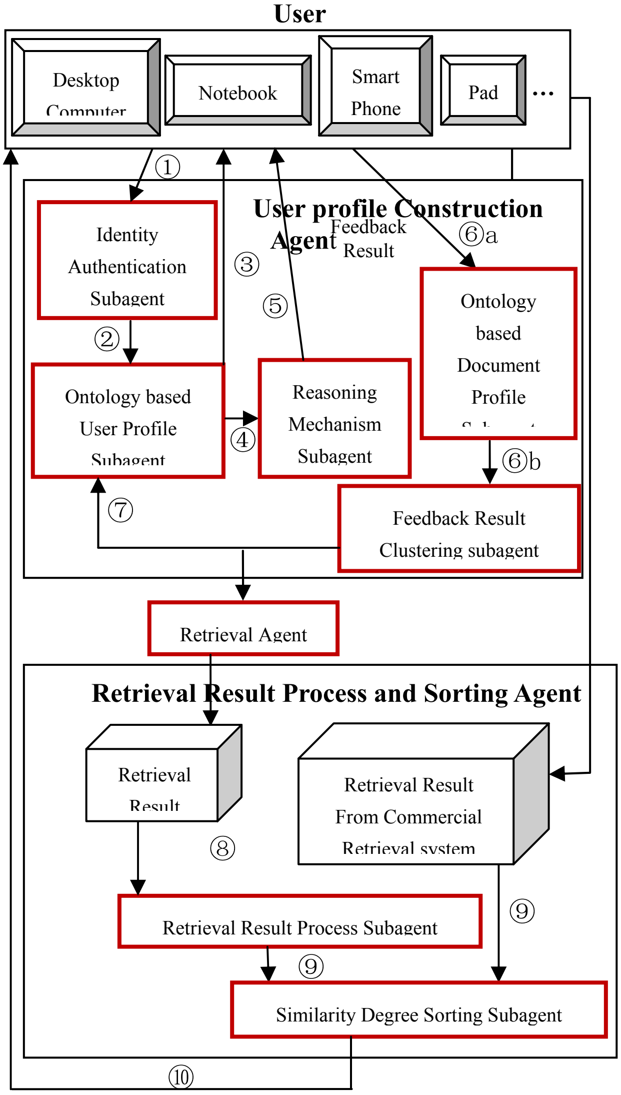
User
Desktop Computer
Notebook
Smart Phone
Pad
…
①
User profile Construction Agent
Feedback Result
⑥a
Identity Authentication Subagent
③
⑤
Ontology based Document Profile
②
Ontology based User Profile Subagent
④
Reasoning Mechanism Subagent
⑥b
⑦
Feedback Result Clustering subagent
Retrieval Agent
Retrieval Result Process and Sorting Agent
Retrieval Result
Retrieval Result From Commercial Retrieval system
⑧
⑨
Retrieval Result Process Subagent
⑨
Similarity Degree Sorting Subagent
⑩

Fig. 1. Framework of the Multi-Agent Context-Awareness User Profile based Personalized Information Retrieval System

Step1: A user i submits a query in the form of "UseID.Query" to the User profile construction Agent, such as User001.news (here "001" is the ID of the user, in order to ensure that even though the user adopts different devices to access our retrieval system, we can identify the user and update the user profile based on their search action; "news" is the query.) (shown as route ① in Fig.1). Then the Identity Authentication Subagent confirms the user's identity and calls the corresponding user profile which stores the user's personal profile according to the user's browning history and interests (shown as route ② in Fig.1). The user's past search experience similar to the query requirement of this time will be found and returned back to the user (shown as route ③ in Fig.1).

Step2: If there is no profile relevant to the query requirement, or the user inputs a fuzzy query, the Ontology-based User profile Subagent will inform the Reasoning Mechanism Subagent (shown as route ④ in Fig.1), and the Reasoning Mechanism Subagent will infer the user's possible demand according to the user's background and feed back to the user (shown as route ⑤ in Fig.1). This process is probably an iterative process

Step3: The ontology-based document profile subagent constructs the document profile in accordance with the user's feedback and the corresponding domain ontology (shown as route ⑥a in Fig.1). Then the feedback result clustering subagent will compare the document profile (shown as route ⑥b in Fig.1) with the relevant interest-domain which has been analyzed according to the ontology-based user profile before the retrieval. This process will determine whether the user's documents of interest belong to the interest-domain, based on which it updates the user profile (shown as route ⑦ in Fig.1)

Step4: The Retrieval Result Process Agent will process the retrieval result such as the removal of redundant, duplicate data (shown as route ⑧ in Fig.1) The Similarity Degree Sorting Subagent will then sort the result according to the similarity matching degree with the input query. Since the creation of the user profiling is a gradual process, the Similarity Degree Sorting Subagent combines our proposed system and the existing search engine with a dynamic weight value α which depends on the perfection degree of our proposed user profile to get higher precision ratio (shown as route ⑨ in Fig.1).

Step6: The Retrieval Result Process and Sorting Agent will finally submit the sorted retrieval results back to the user, and collect the user's feedback result to further cluster the user's interest, and modify the document profile and user profile. (shown as route ⑩ in Fig.1).

3.2. Realization of User Profile Construction Agent

3.2.1. Concept Representation—Domain Ontology

In general, an ontology is the description of a given domain, its classes (or concept), and properties in machine-readable form by means of an ontology language commonly referred to as a knowledge representation model.

Consider that for each domain, most of the concepts are long standing and fixed in spite of some newly born instant concepts, in this paper, for each kind of domain, we use a two-layer model to construct the domain ontology. The first layer is the term ontology O_T, and the other one is the instance layer O_S (For example: companies, products, managers and website are the instances in business domain ontology). The first layer can only be a fixed layer when newly created domain concepts require modification, whereas the other one is a dynamic layer using statistical techniques to update, shown as Fig.2.

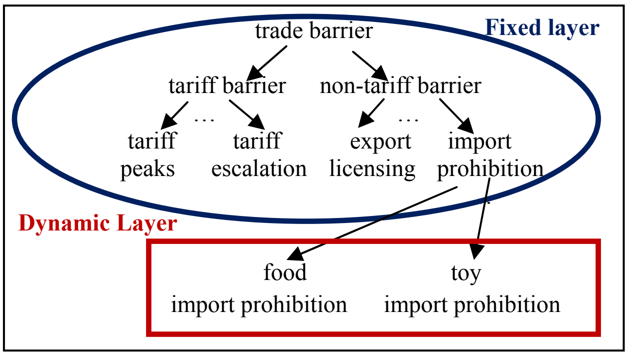

Fig.2. A part of business domain ontology

3.2.2 Ontology based Document Profile Subagent

In order to obtain the users' interests and realize the high precision personalized information retrieval, we adopt an ontology-based knowledge model to create the document profile and user profile.

1) First, find the candidate concept that a document may have in certain domain ontology.

We use the traditional vector space model (VSM) to represent each concept in the domain ontology and the concepts appearing in the document. $V_L(C_i)=\{c_1,c_2,\ldots,c_M\}$ is a one-to-many mapping describing all the text strings in language L which represents concept C_i, where $C_i \in C$, C is the concept of a certain domain ontology.

For each document, $V_d=\{C_{d1},C_{d2},\ldots,C_{dn}\}$is the keyword vector of document d. Suppose $V_\Sigma(D)$ is all the concept collection for a certain domain within document d, $V_\Sigma(D)=\{C_i \in C|\ V_L(C_i)\cap V_d \neq \Phi\}$. We use the "Similarity Matching Algorithm" proposed by Gao & Cho [10] to look up the top N concepts $V_{\Sigma N}(D)$ which have the closest similarity between V_d and $V_L(C)$. Here the value of N is selected by experiment simulation.

2) Second, compute how specific a concept will be.

Different concepts in $V_{\Sigma N}(D)$ have different effects on the document, so that in the document profile, we should assign each concept a relevant weight. Instead of using tf-idf method, we use an ontology-based tf-idf (O-tf-idf)value to determine the importance of the concept in document d.

The ontology-based tf-idf (O-tf-idf) value not only considers the frequency and the specific of the concept in one document, but also considers the generality of the concept in the corresponding domain ontology. The higher the ontology based o-tf-idf value is, the more specific the concept will be.

For a certain concept, more subclasses or more equivalent classes it has in the corresponding domain ontology, more general it will be, so we use a recursive equation (1) to estimate the generality of a certain concept by counting the number of subclass or equivalent class of the corresponding domain ontology.

$$\begin{cases} N(SC_i)=0 \quad SC_i\ is\ the\,leaf\ \ node\ \ in\ \ corresponding\ \ ontology\ \ tree \\ N(SC_i)\geq 1\ \ and\ \ N(SC_i)=\sum\limits_{ssc_i\in subclass\ \ of\,SC_i} N(SSC_i) \qquad otherwise \\ N(C_i)=N(SC_i)+N(EC_i) \quad C_i\in V_{\sum N}(D) \end{cases} \tag{1}$$

Here SC$_i$ is the subclass of the concept C$_i$ and EC$_i$ is the concept that has the similar meaning with C$_i$.

Furthermore, since for a certain concept, different positions it appears in denote the different in significance in the corresponding domain, we take the position of the keywords into consideration. If the keywords that belong to V$_{\sum N}$(D) appear in the title of the first level page, then the weight of this keyword should be higher than the second level page, third level page and so on(Here the first level means the first page of the retrieval results submitted to the user by our system, the second level means the hyperlink of the first page and the third level means the Web page text).

In this paper, we use three weighting parameters K1, K2, K3 (K1>K2>K3) to indicate the importance of the keywords according to their position. Here the value of the K1, K2, and K3 are selected by experiment simulation. We use equation (2) to calculate position value of the keywords.

$$PV(C_i)=\frac{N_1(C_i)\times K1+N_2(C_i)\times K2+N_3(C_i)\times K3}{N_1(C_i)+N_2(C_i)+N_3(C_i)} \tag{2}$$

Here N$_i$ is the occurrence number of the keyword C$_i$ in the ith level. Then, we use equation (3) to calculate the weight of each concept appearing in document d, so as to complete the construction of the document profile.

$$W(C_i)=\beta\times tf_{c,d}\times idf_{c,d}\times\ln\left(\frac{average\left(\sum\limits_{C_i\in V_{\sum N}(D)} N(C_i)\right)}{N(C_i)}\right)\times PV(C_i) \tag{3}$$

Here $tf_{c,d}$ is the number of times concept c occurs in document d, $dfc,_d$ is the number of documents the concept c occurs at least once and the i$dfc,_d$ is the inverse document frequency which can be calculated by equation (4) and β is a normalization constant.

$$idf_{c,d}=\log\frac{|d|}{df_{c,d}} \tag{4}$$

Therefore, the documents can be represented as DVP$_{\sum N}$(D$_i$,W)= DVP$_{\sum N}$ {(C$_1$, W$_1$), (C$_2$, W$_2$),…, (C$_N$, W$_N$)}, where C$_i$∈V$_{\sum}$(D), D$_i$∈the documents that the user feeds back, W$_i$ is the weight of the keyword.

3.2.3. Feedback Result Clustering Subagent

The feedback information means the user's reaction of the retrieval result proposed by our system, including which WebPages are browsed by the user, whether a hyperlink of the WebPages is clicked, browsing time and so on.

The feedback result clustering subagent classifies the user's feedback documents represented by the document profile into a certain interest domain and compares them with all of the interest domains in the user profile, so as to determine whether the feedback documents belong to an existing interest domain in the user profile. If they belong to an existing interest domain in the user profile, then update the user profile based on the feedback documents; otherwise create a new interest domain in the user profile.

Since the user's feedback information is not always an accurate one and the selection of the user probably migrates, so in this paper we use Fuzzy c-means clustering method to cluster the user's interest domain.

$V_{\sum N}(D)=\{C_1,C_2,...,C_N\}\subseteq R^N$ is one of the user's interest documents, which is a finite data set in the characteristic vector space. Here C$_i$ is the characteristic keyword vector. Suppose NC is the number of the interest domain classifications that the document may belong to according to the characteristic keyword, 2≤NC≤N; R^{NC} is an interest domain vector which indicates all the probable interest domains the document may belong to. R^{NNC} is the set of all the N × NC matrixes indicating all the prospective interest domains the document may belong to and the corresponding characteristic keyword of each prospective interest domain.

We use Fuzzy c-means clustering method to get the fuzzy clustering center of the document's each prospective interest domain, shown as equation (5).

$$J=\sum_{i=1}^{N}\sum_{j=1}^{NC}\|c_j^i-ct_i\|^2 \tag{5}$$

Where $c_j^i\in R_j^{NC}$, ct_i is the clustering center of the cluster R_j^{NC}. Then we use fuzzy clustering algorithm to find the fuzzy cluster center:

(1) Assign an initial cluster center for each cluster. So NC numbers of clusters have NC number of cluster centers.

(2) Allocate the characteristic keywords into one of the NC numbers of clusters in accordance with the minimum distance principled $D_i=\min\{\|c-ct_i)\|\},\ \ c\in R^{NC}$, i=1,2,…,NC.

(3) Use the mean of all the characteristic keywords in each cluster as the new cluster center.

(4) If the cluster center changes, then repeat the step (2) and step (3), until the cluster center no longer changes. This is the final cluster center of the NC numbers of clusters.

3.2.4 Ontology based User Profile Subagent

Since each user is likely to be interested in a variety of topics prone to drift over time as new information is available, it is necessary to find a dynamic updating policy to consider the change of the users' preferences over time and domain.

In this paper, we use the method of ontology-based two-level dynamic update algorithm with fuzzy c-means clustering method and dynamic updating policy to realize the construction and updating of the user profile.

1) Ontology-based two-level User Profile presentation

Assume that each user i has its own concept collection $UC_{\sum}(U_i,D_j)$ in domain D_j, which is a subset of the corresponding domain ontology $O(D_j)$. Therefore, the aim of our proposed method is to assign an active weight to each concept belonging to $UC_{\sum}(U_i, D_j)$, so as to reflect the users' interests on this concept.

Here we split the user profile into two components: the fixed concept component which is constant for different users and the dynamic event component, which is dependent on different users. The user profile of a certain user can be represented as n ×m matrixes; each record can be represented as $UCP_{\sum}(U,Di)$= $UCP_{\sum}\{(C_1, W_1), (C_2, W_2),\ldots, (C_N, W_N)\}$, where $C_i \in UC_{\sum}(U, D_i)$, W_i is the weight of the keyword, and each field indicates a user.

In the beginning, we initiate the weight of the concepts that belong to $UC_{\sum}(U_i, D_j)$ using the weight $WU(C_i)$, as shown by equation(6).

$$\begin{cases} WU(C_i)=W(C_i) & C_i \in \ V_{\Sigma N}(D_i)\, and \quad C_i \in C_{\Xi}(U_i,D_j) \\ WU(C_i)=0 & C_i \in (C_{\Sigma}(U_i,D_j) \quad but \notin V_{\Sigma N}(D_i) \end{cases} \qquad (6)$$

2) Interval value fuzzy set mapping of cluster center and user's own concept collection node.

With the method proposed in section 3.2.3, we have obtained the user's behaviors and formed the cluster. But since it is difficult for the cluster problem to have an exact value, in this paper, we adopt an interval value fuzzy set to indicate the center of the clustering result, shown as equation (7).

$$OC(x)=[OC^{-}(x),OC^{+}(x)] \qquad x\in R^{NC} \qquad (7)$$

We also use the interval value fuzzy set to express the node in the user's own concept collection, shown as equation (8).

$$UC(x_i)=[UC^{-}(x_i),UC^{+}(x_i)] \qquad x\in C_{\Sigma} \qquad (8)$$

Then we use equation (9) which can evaluate the intersection degree of two fuzzy sets to find the similar part of the user's own concept collection with the center of the clustering result.

$$S=[\vee_{x\in R^{NC}}\{OC^{-}(x)\wedge UC^{-}(x)\},\vee_{x\in C_{\Sigma}}\{OC^{-}(x)\wedge UC^{+}(x)\}] \qquad (9)$$

If S=[1,1], it means that the two fuzzy sets intersect, otherwise if S=[0,0], it means that the two fuzzy sets have no intersection.

3) Dynamic updating strategy.

Suppose the users' own concept collection is also represented as ontology tree, called user personalized ontology. For one concept in the user personalized ontology, it may also have equality class and subclass. Furthermore, the more frequency the concept appears in the users' positive feedback documents the higher the weight of this concept will be, and if the weight of the concept is higher than a threshold, the concept should be split into two or more subclass concepts; on the contrary, if the weight of the concept is lower than a threshold, the concept should be eliminated or be merged into its parent concept.

Therefore, for all of the intersections of two fuzzy sets, we comprehensively consider the generality of the concept, the number of the newly inserted positive feedback documents, and the similarity between the generality of the concept and the weight of the concept in current documents' profile to update the weight of the related concept in the users' profile, shown as equation (10).

$$WU(C_i)_k=\alpha\times\sum_{k\in D_j}N(C_k)\times\frac{1}{|D_j|}+(1-\alpha)\times\sum_{k\in D_j}\cos(WU(C_i)_{k-1},N(C_k)) \qquad (10)$$

Where D_j is the set of the newly inserted documents of concept in period j. | D_j | is the number of the documents in D_j. C_k is the concept vector in the profile of document k. $Cos(WU(C_i)_{k-1},N(C_k))$ is the cosine similarity between the weight of the concept C_i in current user profile and the generality of the concept C_k which appears in the newly inserted documents in period j.

4) Fuzzy c-means clustering and dynamic updating based user profile updating algorithm

```
Input: Document profile $WVP_{\sum N}(D_i,W)$, Initial user profile $UCP_{\sum}(U,Di)$.
Output: Modified user profile
Step1: Call the fuzzy clustering algorithm to generate the fuzzy clustering center $D_c$.
Step2: Compare the $D_c$ with the user personalized ontology D to find the sub domain $D_j$ the $D_c$ belongs to.
Stpe3: For i=1 to n do      //n is the node number of the sub domain $D_j$
          Use equation (9) to compare the two fuzzy sets
          If m numbers of S≥preset threshold P then
               Add $D_c$ into the all the $D_j$ (j=1,2,…,m) and use equation (10) to update the w  eight of the concepts in domain $D_j$ (j=1,2,…,m).
          Else      // no S≥preset threshold P
               Add $D_c$ as a new node into the user personalized ontology D, and use the weight of $D_c$ as the weight of the new born node.
Step4: End.
```

3.3. Realization of Retrieval Result Process and Sorting Agent

3.3.1. Retrieval Result Process Subagent

The creation of the user profiling is a gradual process, so that in the initial, we could not only rely on our user profiling-based IR system to get the search result. Thus, we combine our proposed system and the existing search engine with a dynamic weight value δ.

The value of δ depends on the completeness of the user profiles. At first, we assign a lower value to δ, and with the completeness of the user profile, δ will increase. Here we consider the top 20 results. The value δ means that from 20 results, our system selects top δ results from user profile search and selects top 1-δ results from an existing search engine, shown as Fig.3.

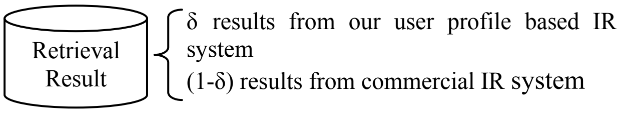
Retrieval Result
δ results from our user profile based IR system
(1-δ) results from commercial IR system

Fig.3. Retrieval result selecting and page ranking strategy

At the beginning, since the user profile is not perfect, the value of δ should decrease in the former, but along with the gradual perfection of the user profile, the value of δ should increase gradually. Here we adopt equation (11) to modify the value of δ according to the click rate of our method and the commercial information retrieval system, such as Naver.

$$\begin{cases} \delta = e^{c} - 1 & c \le 0.5 \\ \delta = (c-0.5)^{\frac{1}{3}} + 0.2 & c > 0/5 \end{cases} \tag{11}$$

Where the click rate c=(Clicked number from our method result)/(Total click number). Furthermore, in order to avoid the effect of the sudden click, we impose the restrictions as follows: if |C-δ|>0.6 then the value of δ will not be modified.

3.3.2 Similarity Degree Sorting Subagent

Suppose the query input by the user is Q={q$_1$,q$_2$,…,q$_n$}, we use equation (12) to evaluate the similarity of the query and the retrieval documents, based on which to sort the retrieval documents.

$$Sim(D,Q) = \sum_{i=1}^{n} \ln \frac{\dfrac{N_{rdq_i}}{N_{rd}}\left(1 - \dfrac{N_{dq_i} - N_{rdq_i}}{N_d - N_{rd}}\right)}{\dfrac{N_{dq_i} - N_{rdq_i}}{N_d - N_{rd}}\left(1 - \dfrac{N_{rdq_i}}{N_{rd}}\right)} \tag{12}$$

Here, N_d means the total number of the training documents; N_{rd} means that the documents are relevant to the user's query; N_{rdq_i} means that the documents are relevant to the user's query and include term q$_i$, N_{dq_i} means the number of the training documents containing the tern q$_i$.

4. Simulation

We construct a HDFS Cluster to simulate our proposed method (a single computer with an Intel Core2 CPU, 4 GB RAM as NameNode, and 5 computers with Intel Core2 CPUs and 4 GB RAM as DateNode). We use Java as the programming language and the Eclipse integrated development environment. In this study, we analyze and operate the ontology files with an open-source framework "Jena."

To test the approach proposed in this paper, we adopt the TREC 10 query sets as the tools, and eight Ph.D. and Masters students from different fields of computer science and education as the participants in the experiments.

We compare our system results with the semantic based technique [11] which uses traditional vector space model (VSM), also called term frequency (tf) inverse document frequency (idf) to carry out searching, and evaluate the performance of the two methods from the precision and recall. The Precision (P) is the fraction of retrieved documents relevant which is defined as Equation (13). Recall (R) is the fraction of relevant documents retrieved which is defined as Equation (14)

Precision =#(relevant items retrieved)/#(retrieved items)
=P(relevant|retrieved) (13)

Recall =#(relevant items retrieved)/#(relevant items)
=P(retrieved|relevant) (14)

The simulation results are shown in Figs. 4and 5.

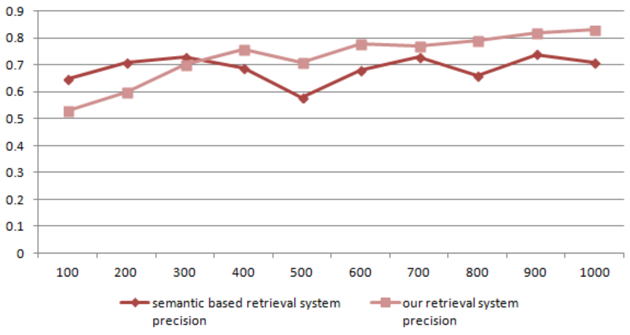

Fig. 4. Precision ratio of semantic based retrieval system versus our proposed retrieval system for TREC 10

The overall results show that at the beginning, since the user profile is not perfect, more retrieval results will be selected by the commercial information retrieval system and the Precision Ratio and the Recall Ratio of our proposed system is lower than the semantic based retrieval system. Along with the gradual perfection of the user profile, more retrieval results will be selected by our system and more users' interests will be considered, so our approach is better than semantic based retrieval system in terms of the Recall Ratio and Precision Ratio. In addition, since the users' long-term interests may change with time and the short-term interests may occur (maybe only for once retrieval), there might be a fluctuation on Recall Ratio and Precision Ratio before the users' profile is perfect. But once the user profile is perfect, the Recall Ratio and Precision Ratio will be maintained on a steady level

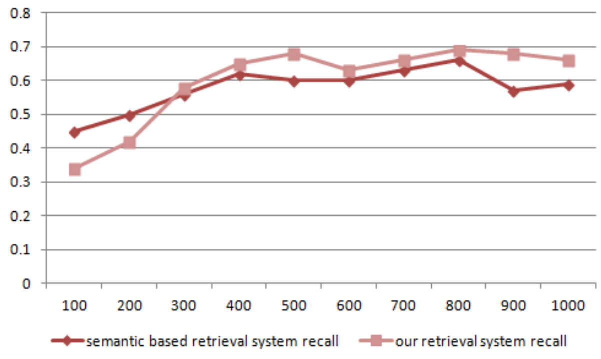

Fig. 5. Recall ratio of semantic based retrieval system versus our proposed retrieval system for TREC 10

5. Conclusion

This paper attempts to create a multi-agent context-awareness method based on dynamic ontology that incorporates concepts and properties to model the user profile. The novelty of this paper is that it adopts a two-layer (term layer and instance layer) model to construct the domain ontology and uses ontology-based o-tf-idf method to calculate the weight of each concept which comprehensively considers the frequency and the specific of the concept in one document and its corresponding domain ontology. Another novelty is that this paper adopts a fuzzy c-means clustering method to cluster the user's interest domain, which is the basis to update the users' profile by using a dynamic updating strategy, so as to continuously track the users' current interests. Furthermore, in order to meet the users' requirements to maximum, we comprehensively use our proposed retrieval system and commercial retrieval system which can avoid the lower precision in the beginning caused by the imperfective user profile.

The simulation result shows that along with the gradual perfection of the our user profile, our proposed system is better than semantic based retrieval system in terms of the Recall Ratio and Precision Ratio.

In the future, more researches should be done about how to combine the collaborative learning algorithm with our system, so as to reduce the number of preference request calls when a large database of users is available or new topics are created.

References

[1] J. Wang, Z. Li, J. Yao, Z. Sun, M. Li, and W. Ma, "Adaptive User Profile Model and Collaborative Filtering for Personalized News," *Proc. of the APWeb 2006*, pp. 474-485, 2006.

[2] L. Suryanarayana and J. Hjelm, "Profiles for the situated web," *Proc. of the 11th international conference on World Wide Web*, pp.200-209, 2002.

[3] J. Trajkova and S. Gauch, "Improving Ontology-based User Profiles," *Proc. of RIAO 2004*, pp. 380-389, 2004.

[4] S. Gauch, J. Chaffee, and A. Pretschner, "Ontology-Based User Profiles for Search and Browsing," *Web Intelligence and Agent Systems*, vol. 1, no. 3-4, pp. 219-234, 2003.

[5] A. Maedche and S. Staab, "Mining Ontologies from Text," *EKAW 2000*, pp. 189-202, 2000.

[6] L.Aroyo, R. Denaux, V. Dimitrova, and M. Pye, "Interactiv e Ontology based User Knowledge Acquisition: A Case Stu dy," *European Semantic Web Conference*, pp. 560-574, 2006.

[7] H. Zhang and Y. Song, "Construction of Ontology-based User Model for Web Personalization ," *Lecture Notes in Computer Science*, vol. 4511, pp. 67-76, 2007.

[8] D.H. Kraft, G. Bordogna, and G. Passi, "Information Retrieval Systems: Where is the Fuzzy?," *Proc. of IEEE International Conference on Fuzzy Systems*, pp. 1367-1372, 1998,

[9] Ph. Mylonas, D. Vallet, P. Castells, M. Fern´andez, and Y. Avrithis, "Personalized information retrieval based on cont ext and ontological knowledge," *The Knowledge Engineeri ng Review*, vol. 23, Issue 1, pp. 73-100, 2008.

[10] Q. Gao and Y. I. Cho, "Similarity Matching Algorithm for Ontology-Based Semantic Information Retrieval Model," *Proc. of the 10th World Congress on Intelligent Control and Automation*, pp. 758-763, 2012.

[11] D. Bonino, F. Corno, L. Farinetti, and A. Bosca, "Ontology Driven Semantic Search," *WSEAS Transaction on Information Science and Application*, Issue 6, vol. 1, pp. 1597-1605, 2004.

Qian Gao

2001 B.A. Shandong University of Science and Technology, China

2008 Master Shandong University.

2012 current Ph.D Course student.

Interest areas: neural network, fuzzy logic, artificial intelligence, information retrieval, ubiquitous computing, etc.

Young Im Cho

1994 Ph.D Dept. of Computer Science, Korea University, Korea

1996 Senior Researcher of Samsung electronic.

2012 current Professor of the University of Suwon.

Interest areas: Intelligent system, ubiquitous system, information retrieval, neural network, fuzzy system, etc.