Advances in Engineering Software 40 (2009) 535–540

Contents lists available at ScienceDirect

Advances in Engineering Software

journal homepage: www.elsevier.com/locate/advengsoft

A fuzzy clustering method of construction of ontology-based user profiles

Lixin Han [a,b,c,*], Guihai Chen [b]

[a] College of Computer and Information Engineering, Hohai University, China
[b] State Key Laboratory of Novel Software Technology, Nanjing University, Nanjing, China
[c] Department of Mathematics, Nanjing University, Nanjing, China

ARTICLE INFO

Article history:
Received 15 August 2008
Accepted 21 October 2008
Available online 9 December 2008

Keywords:
Fuzzy clustering
User profile
Optimization

ABSTRACT

With the rapid development of personalized information retrieval, user profile plays an important role. In this paper, we propose a fuzzy clustering method of construction of ontology-based user profiles (FCOU). In the FCOU method, we employ fuzzy clustering techniques combined with optimization techniques to develop ontology-based user profiles. One key feature of FCOU is that it employs an augmented Lagrangian function to create fuzzy clustering model for the construction of user profiles. Another key feature of FCOU is that it employs the combination of FCM, PHR and simulated annealing to develop ontology-based user profiles. The method allows some information to belong to several user profiles simultaneously with different degrees of accuracy.

1. Introduction

Nowadays the amount of information on the Internet is increasing dramatically. The ability of facilitating users to achieve useful information is more and more important for information retrieval system. A user profile is required to present the information that the user truly needs.

Generally, there are two major approaches for collecting user profiles [1]. Preprocessing is a method done with a questionnaire. This method is time consuming for users. For the other method, automatic collection is based on the interaction between the user and the system. User profiles are acquired by analyzing the web pages that the user visits. For example a user profile is created by analyzing the content of surfed pages and even by analyzing the length of the document and the time that was spent on the document associated with that content. The cost of automatic collection is not higher than the cost of preprocessing. However, when systems use the acquired information only for their own system, the difference of the accuracy of adaptation may occur. For example, when the user frequently uses a system, the system can achieve an adaptation very well. However, when the user rarely uses a system, the system cannot achieve an adaptation very well. To solve this problem, the user must repeat a similar procedure for each system. This is a burden to the user.

We propose a method called FCOU (fuzzy clustering method of construction of ontology-based user profiles). The FCOU method is used to develop ontology-based user profiles. The user profiles can maintain sophisticated representations of personal interest profiles. These representations can be utilized for effective information retrieval. Fuzzy clustering allows an entity to belong to more than one cluster with different degrees of accuracy, while hard clustering assigns each entity exactly to one of the clusters. Thus, fuzzy clustering is suitable in constructing ontology-based user profiles because some information is not forced to fully belong to any one of the user profiles. Fuzzy clustering methods may allow some information to belong to several user profiles simultaneously with different degrees of accuracy. FCOU is different from both classes of the above methods in that our algorithm employs the combination of fuzzy clustering and optimization to construct user profiles.

2. Related work

The fuzzy c-means (FCM) algorithm [2] is one of the best known methods in fuzzy clustering. FCM introduces the concepts of fuzzy logic to classic K-means. Based on FCM, the fuzzy clustering multiple prototype (FCMP) framework [3] proposes a model of how the data are generated from a cluster structure to be identified. Nascimento et al. [4] extend the FCMP framework to some clustering criteria, and study the FCMP properties on fitting the underlying proposed model from which data is generated.

Tjhi and Chen [5] propose an algorithm called FCC-STF for clustering standard text documents, in order to expand the existing fuzzy clustering algorithms such as Fuzzy CoDoK and FCCM. The FCC-STF algorithm is different from the above two algorithms in that CC-STF uses a different fuzzifier.

Ahn et al. [6] apply open user models to adaptive news systems, in order that the adaptive system becomes more transparent and controllable to the user. They explore the role of open and editable

* Corresponding author. Address: College of Computer and Information Engineering, Hohai University, China.
E-mail address: lixinhan2002@yahoo.com.cn (L. Han).

0965-9978/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.advengsoft.2008.10.006

user profiles so that the users can see view and edit their interest profiles right after changes are made.

Hu et al. [7] present a method of predicting users' gender and age based on their Web browsing behaviors. The solution consists of three steps. In Step 1, users' known Webpage's gender and age tendency is learned using the linear form of Support Vector Machine (SVM) Regression. In Step 2, users' unknown age and gender are predicted using a Bayesian framework. To solve the data-sparseness problem suffered in the above two steps, Singular Value Decomposition (SVD) is used to produce a low-dimensional representation of original user-page space by making use of similarity relationship between users and Webpages.

Nébel et al. [8] propose a method for the creation of user-profiles through the construction of a central database resource of user ontologies, from which ready-made taxonomies of different types of users can be extracted as a unit. The advantages of the method are its easy adaptability, high speed of operation, automatic completion of inferences to yield new information, and automatic extendibility.

Razmerita et al. [9] address the aspects of ontology-based user modeling and they present a generic architecture for implicit and explicit ontology based-user modeling. They identify the aspects of user modeling relevant to KMSs and integrate them in a generic framework based on ontologies.

Kanjo et al. [1] propose A3, which is a framework for user adaptation. In this framework, the automatic construction of the user ontology and the sharing of the user ontology are enabled in achieving user adaptation. The user ontology is defined as a classified tree of web resources, and is written by RDF(S) and OWL proposed for the Semantic Web. They assume that the user ontology represents the user's knowledge, and they use the user ontology as a user profile to achieve a user adaptation on A3.

Kim et al. [10] present their development of a document management and retrieval tool, which is named Ontalk. Ontalk provides a semi-automatic metadata generator and an ontology-based search engine for electronic documents. Ontalk can create or import various ontologies in RDFS or OWL for describing the metadata. Their system that is built upon .NET technology is easily communicated with or flexibly plugged into many different programs.

Nakaya et al. [11] describe how to exploit a machine-readable dictionary (MRD) and domain-specific text corpus in supporting the construction of domain ontologies that specify taxonomic and non-taxonomic relationships among given domain concepts. In order to acquire taxonomic relationships, they use matching result analysis and trimmed result analysis. In order to learn non-taxonomic relationships, they use WordSpace and an association rule algorithm.

Maedche et al. [12] discuss how ontology learning greatly facilitates the construction of ontologies by the ontology engineer. The framework of ontology learning proceeds through ontology import, extraction, pruning, refinement, and evaluation. Besides of the general framework and architecture, they show some exemplary techniques in the ontology learning cycle and refer to some others that need to complement the complete architecture.

Li et al. [13] employ association sets to automatically discover ontologies from data sets. They also set up a reasoning model for capturing evolving patterns in the ontology, in order to refine its association set. In addition, they present a formal method of automatically learning how to assess relevance in the ontology.

Bouquet et al. [14] present a method which can be used to automatically elicit and represent the meaning of a schema from very common web objects. This method takes into consideration not only structural knowledge but also lexical and domain knowledge about the labels used in the schema. In this method, meanings are represented in a formal language called WDL (wordnet description logic) which is the result of combining a logical language ALCIO and a well-known electronic lexical database WORDNET.

Suchanek [15] present a large and extendable ontology called YAGO. In an ontology, the Is-A hierarchy and non-taxonomic relations between entities are built. It consists of not only individuals from Wikipedia but also the taxonomy of concepts from WordNet. They employ a combination of rule-based and heuristic methods to extract the facts from Wikipedia and WordNet.

Trajkova and Gauch [16] propose a method of the construction of user profile based on the concepts from a pre-existing ontology. The method represents the user profile by the total weight and number of Web pages that are associated with each concept in the ontology, in order to classify the collected Web pages into the top-matching concepts in the predefined ontology using the vector space model. In addition, this paper report how to improve the accuracy of the profile generated.

In contrast to the above related work, the FCOU method employs an augmented Lagrangian function to create fuzzy clustering model for the construction of user profiles. The FCOU method employs the combination of FCM, PHR algorithm and simulated annealing to develop ontology-based user profiles.

3. The FCOU method

In the FCOU method, an augmented Lagrangian function is used to create fuzzy clustering model for the construction of user profiles. The OSGP algorithm is then proposed to compute the optimal solution under given a vector of Lagrange multipliers estimates and a penalty parameter conditions. Based on the OSGP algorithm, the CPO algorithm is proposed to compute the optimal solution. Based on the CPO algorithm, the CCF algorithm is proposed to construct user profiles.

In FCOU, the ontology-based user profiles are used to express sophisticated user profiles in order that a user may own several interests. A user profile may correspond to the several ontologies. For example, there are two ontology *faculty* and *student*. In ontology *faculty*, faculty is a superclass, positions, research interests and teaching activity are subclasses. At the next lower level, teaching activity is a superclass. Course name and credit are subclasses. In the other ontology *student*, student is a superclass. Undergraduate student and graduate student are subclasses. At the lower level, graduate student is a superclass. Master student and Ph.D. student are subclasses. Xin Chen's user profile is described as follows: {Xin Chen, professor, (information retrieval, database, semantic Web) (Jue Zhang, Hao Guo, Yin Zhao, Ding Hu, Ling Hua)}, {(information retrieval, 2)}}. Ning Huang's user profile is described as follows: {Ning Huang, assistant professor, (database, semantic Web) (Hao Guo, Ling Hua), {(database, 3), (semantic Web, 2)}}, where {denote layer. Xin Chen and Ning Huan's user profiles are explained below: Xin Chen is a professor. He teaches information retrieval that carries 2 credits. His research interests are information retrieval, database and semantic Web. Ning Huang is an assistant professor. He teaches database that carries 3 credits and semantic Web that carries 2 credits. His research interests are database and semantic Web. Ph.D. student Jue Zhang, Yin Zhao and Ding Hu are supervised under the same advisor Xin Chen. Ph.D. student Hao Guo and Ling Hua are co-supervised under the advisor Xin Chen and Ning Huang. Therefore Hao Guo and Ling Hua are not forced to fully belong to anyone of Xin Chen and Ning Huan.

3.1. The fuzzy clustering model for construction of ontology-based user profiles

The most known method of fuzzy clustering is the fuzzy *c*-means method (FCM) [2]. FCM introduces the concept of fuzzy sets to classic K-means [17,18]. Fuzzy set theory is an extension of the classic set theory developed by Zadeh [19] as a way to deal with vague concepts. Classical set theory considers an object as a member of a given set or not, that is, indicator variable is 1 and 0. In a fuzzy set, the indicator variable called membership can take intermediate values in interval [0,1].

The FCM algorithm assumes that the number of clusters c is known in advance and minimizes an objective function to find the best set of clusters. Usually, membership functions are defined based on a distance function, such that membership degrees express proximities of entities to cluster prototypes.

In the FCM algorithm, let $X = \{x_1, \ldots, x_n\}$ denote a set of unlabeled feature vector in R^p, and let c be an integer, $1 < c < n$. Each x_j is the numerical representation of p features. Given X, a fuzzy c-partition of X is represented by a $c \times n$ fuzzy partition matrix $U = [u_{ij}]$ satisfying the conditions: $0 \leqslant u_{ij} \leqslant 1$ $(1 \leqslant i \leqslant c,\ 1 \leqslant j \leqslant n)$, $\sum_{i=1}^{c} u_{ij} = 1$ $(1 \leqslant j \leqslant n)$ and $\sum_{j=1}^{n} u_{ij} > 0$ $(1 \leqslant i \leqslant c)$, where each value u_{ij} represents the membership of the jth feature vector to the ith cluster. The clustering criterion used by the FCM algorithm is associated with the generalized least-squared errors function.

$$
\begin{aligned}
\min \quad & J_m(U,V) = \sum_{i=1}^{c}\sum_{j=1}^{n}(u_{ij})^m D_{ij} \\
\text{s.t.} \quad & \sum_{i=1}^{c} u_{ij} = 1, \quad \forall j \in \{1,\ldots,n\} \\
& 0 \leqslant u_{ij} \quad i \in \{1,\ldots,c\},\ j \in \{1,\ldots,n\}
\end{aligned}
\tag{1}
$$

where c is the number of fuzzy clusters, $u_{ik} \in [0,1]$ is the degree of membership of feature point x_k in cluster i. Parameter $m > 1$ is the degree of fuzzification called fuzzifier in order to increase or decrease the fuzziness. Higher values of fuzziness will make the result fuzzier. $U = [u_{ij}]$ is a $c \times n$ constrained fuzzy c-partition matrix. If $m \to 1$, then the membership degrees $u_{ik} \to 0/1$, so the classification tends to be crisp. If $m \to \infty$, then $u_{ik} \to \frac{1}{c}$, where c is the number of clusters. $V = [v_1 \cdots v_c]$ $(v_i \in R^p)$ is the vector of cluster prototypes, and D_{ij} is some distance metric between feature vector x_j and cluster prototype v_i, which is taken equal to the squared distance

$$D_{ij} = \|x_j - v_i\|_A^2 = (x_j - v_i)^{\mathrm{T}} A (x_j - v_i) \tag{2}$$

where the matrix A represents a positive definite $n \times n$ weight matrix. If A is taken as the identity matrix I, the resulting Euclidean norm implies hyperspherical clusters.

To minimize criterion J_m, under the fuzzy constraints $\sum_{i=1}^{c} u_{ij} = 1$ and $0 \leqslant u_{ij}$, we describe an augmented Lagrangian function as follows:

$$
\begin{aligned}
J_m(U,V,\lambda,\delta) = & \sum_{i=1}^{c}\sum_{j=1}^{n}(u_{ij})^m D_{ij} - \lambda_1\left(um_{i=1}^{c} u_{ij} - 1\right) + \frac{\sigma_1}{2}\left(\sum_{i=1}^{c} u_{ij} - 1\right)^2 \\
& + \frac{1}{2\sigma_2}\{[\max(0, \lambda_2 - \delta_2 \mu_{ij})]^2 - \lambda_2^2\}
\end{aligned}
\tag{3}
$$

3.2. The OSGP algorithm

Simulated annealing is a generalization of a Monte Carlo method for examining the equations of state and frozen states of n-body systems [20]. The concept is based on the manner in which liquids freeze or metals recrystallize in the process of annealing. By analogy the generalization of this Monte Carlo approach to combinatorial problems is straight forward [21,22]. The major difficulty in implementation of the algorithm is that there is no obvious analogy for the temperature T with respect to a free parameter in the combinatorial problem. Furthermore, avoidance of entrainment in local minima is dependent on the "annealing schedule", the choice of initial temperature, how much iterations are performed at each temperature, and how much the temperature is decremented at each step as cooling proceeds [23].

The OSGP (the optimal solution under given parameters) algorithm which introduces simulated annealing is proposed to compute the optimal solution of the J_m under given a vector of Lagrange multipliers estimates and a penalty parameter conditions. The simulated annealing algorithm can be viewed as globally optimizing an objective function. The OSGP algorithm can ensure to break away current local optimal solution and to reduce the computing workload.

The OSGP algorithm is based on the following solution form:

- Solution space is {feature vector};
- Objective function is

$$
\begin{aligned}
f(U,V,\lambda,\delta) = \min & \sum_{i=1}^{c}\sum_{j=1}^{n}(u_{ij})^m D_{ij} \\
& - \lambda_1\left(\sum_{i=1}^{c} u_{ij} - 1\right) + \frac{\sigma_1}{2}\left(\sum_{i=1}^{c} u_{ij} - 1\right)^2 \\
& + \frac{1}{2\sigma_2}\{[\max(0, \lambda_2 - \delta_2 \mu_{ij})]^2 - \lambda_2^2\};
\end{aligned}
$$

- New solution is produced by randomly choosing another feature vector from neighboring region $N(x)$;
- Objective function difference is $\Delta f = f$(new feature vector) $- f$(local optimal solution);
- Accept criterion is $P = 1$ for $\Delta f > 0$ or $P = \exp(-\Delta f/t)$ for $\Delta f \leqslant 0$.

The OSGP algorithm is described as the following steps:

```
{given initial solution x*, initial optimal objective function value
  f*,
initial temperature t0, initial step q0, monotone increment
  function I(x);
q = q0; // initialize step q
t = t0; // initialize temperature t, t0 = 50
i = 1; // initialization
at = 1; // initialization
while |Ri −C| > ε//ε is a given enough small positive number,
  where ε = 3 × 10^−3. C is an threshold, where C = 0.79
{Ri−1 = at/as; //at is the acceptable times of new solution, as is
  iterative steps number, Ri−1 is acceptable rate
  at = at + 1;
  as = 0; // initialization
  while J<>U //U is the upper limit of step length
  {as = as + 1;
    while x∉N(x) // N(x) is neighboring region
  {step length is q and the BFGS optimizer in the quasi-Newton
    method is used to compute local optimal solution;
  }
xi is produced by randomly choosing another feature vector from
neighboring region N(x) and objective function difference
  Δf = f(xi) − f(x) is computed;
if Δf ⩽ 0 or exp(−Δf/t) > random (0,1)
then {x = xi;
      if f < f*
  then {x = x*;
        f = f*;
        }
  if at is equal to the given acceptable times
  then break;
  }
```

$q = I(R_i) \times q$; // adjustment function $I(x)$ is monofonic increasing function, where $I(x) = (x - 0.5)^5 + 1$
}
if $f < f^*$
then {$x = x^*$;
$f = f^*$;
}
$R_i = at/as$;
if $i > 2$
then // compute self-adjusting temperature
{**if** $R_{i-1} < C$ and $R_i < C$
then $t = t + TC$; // TC is a given constant
if $R_{i-1} \geqslant C$ and $R_i \geqslant C$
then $t = t - TC$;
if $R_{i-1} \leqslant C$ and $R_i \geqslant C$
then $t = t - TC/2$;
if $R_{i-1} > C$ and $R_i < C$
then $t = t + TC/2$;
}
$i = i + 1$;
}
}

In contrast to the widely used simulated annealing algorithm [21,24], there are three features in the OSGP algorithm. Firstly, the OSGP algorithm can self-adaptedly adjust temperature. Therefore it is regarded as time after time optional process. On the one hand, the Boltzmann factor $\exp(-\Delta f/t)$ increases with t. This is useful to break away current local optimal solution and to increase step length in order to look for a better new solution. On the other hand, when acceptable rate meets the given condition, the program can end so as to reduce the computing workload. Secondly, local optimization techniques instead of choosing randomly a new solution are employed in order to compute a local optimal solution. Thirdly, the local optimal solution is recorded during annealing process, in order to protect the current optimal solution from being thrown away. Therefore the OSGP algorithm is an intelligent algorithm.

3.3. The CPO algorithm

The CPO (combines PHR with OSGP) algorithm that combines the PHR algorithm [25] and the OSGP algorithm is used to compute the optimal solution.

The CPO algorithm is described as follows:

{given initial solution x_0, initial a vector of lagrange multipliers estimates λ_1 and λ_2, initial a penalty parameter δ_1 and δ_2, a user-given small precision $\varepsilon > 0$, a amplification coefficient $c > 1$, constant $\theta \in (0,1)$; // initialization
$k = 1$; // initialization
initialize optimal solution x^*;
while $¢^k > \varepsilon$
{the OSGP algorithm is used to compute the optimal solution of the J_m; //given $\lambda_1^k, \lambda_2^k, \delta_1^k, \delta_2^k$
$\varphi^k = \{(\sum_{i=1}^{c} u_{ij} - 1)^2 + \max(\mu_{ij}, \lambda_2^k/\delta_2)\}$;
if $¢^k/¢^{k-1} > \theta$
then {$\delta_1^{k+1} = c\delta_1^k$; // updating penalty parameter δ_1^k
$\delta_2^{k+1} = c\delta_2^k$; // updating penalty parameter δ_1^k
}
else {$\delta_1^{k+1} = \delta_1^k$;
$\delta_2^{k+1} = \delta_2^k$;
}
$\lambda_1^{k+1} = \lambda_1^k - \delta_1^k(\sum_{i=1}^{c} u_{ij} - 1)$; //updating lagrange multipliers estimates λ_1^k
$\lambda_2^{k+1} = \max[(0, \lambda_2^k - \delta_2^k u_{ij})]$; //updating lagrange multipliers estimates λ_2^k
$k = k + 1$;
}
}

In contrast to the PHR algorithm, The CPO algorithm employ simulated annealing to compute the optimal solution of the J_m under given a vector of Lagrange multipliers estimates and a penalty parameter conditions.

3.4. The CCF algorithm

We propose the CCF (combines CPO with FCM) algorithm that combines the CPO algorithm and the FCM algorithm for constructing ontology-based user profiles.

The CCF algorithm is described as follows:

{given a set of unlabeled feature vector X, the number of clusters C, fuzzifier $m > 1$, the maximum number of iterations L and the termination tolerance $\varepsilon > 0$;
initialized $U^{(0)}$ with random numbers, and normalized to make row sums equal to 1;
$l = 0$;
do {$l = l + 1$; // the iteration index l
$v_i^{(l)} = \frac{\sum_{k=1}^{N}(u_{ik}^{l})^m x_k}{\sum_{k=1}^{N}(u_{ik}^{l})^m}$ for $1 \leqslant i \leqslant c$; // V_i is the cluster centers
$D_{ik}^2 = \|x_k - v_i\|^2 = \sqrt{(x_k - v_i)^T A (x_k - v_i)}$ for $1 \leqslant i \leqslant c, 1 \leqslant k \leqslant N$;
// D_{ik} is the distances, A represents a positive definite $n \times n$ weight matrix
if $D_{ik} > 0$ for $1 \leqslant i \leqslant c, 1 \leqslant k \leqslant N$
then $u_{ik}^{(l)} = \frac{1}{\sum_{j=1}^{c}\left(\frac{D_{ik}}{D_{jk}}\right)^{\frac{2}{m-1}}}$;
else $u_{ik}^{(l)} = 0$; //update the partition matrix
the CPO algorithm is used to compute the optimal solution of the J_m;
}
while $\|U^{(l)} - U^{(l-1)}\| \geqslant \xi$ and $l < L$ // membership matrix $U^{(l)}$ with
elements $u_{ki}^{(l)}$, where $0 \leqslant u_{ki}^{(l)} \leqslant 1, \sum_{i=1}^{c} u_{ki}^{(l)} = 1$, for all k
some new rules are generated from these clusters;
these rules are stored into the knowledge base;
the widely used Jena2's inference mechanism [26,27] is used to infer semantic associations from the existing rules in the knowledge base in order to discover semantic relationships among the clusters;
}

In contrast to the widely used the fuzzy c-means (FCM) algorithm [2], the CCF algorithm employs a combination of the PHR and simulated annealing algorithm to compute the optimal solution and provides the inference mechanism to find the semantic relationships among the clusters.

4. Experiment results and discussion

In this section, we conduct the experiments to evaluate the performance of the FCOU method.

4.1. Experimental results and analysis

We applied the FCM algorithm to automatically construct ontology-based user profiles. We randomly select two faculties,

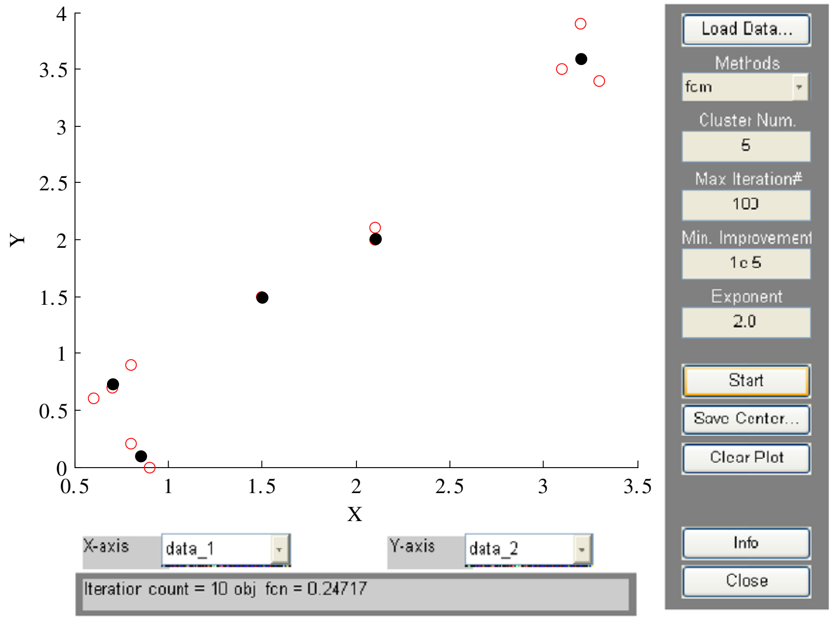

Fig. 1. Experimental result 1.

and their personal information such as positions, research interests, graduate students and teaching activity. The element vector a_{ij} in a two-dimensional vector space represents the *ith* relevant information that appears in the *jth* person. A set of 16 nodes are arranged in a two-dimensional space with weight vector attached to each node. The data is described below:

1.5	1.5
0.8	0.2
0.6	0.6
0.8	0.9
0.7	0.7
0.9	0
2.1	2.0
2.1	2.1
2.1	2.0
2.1	2.0
2.1	2.0
2.1	2.0
2.1	2.0
3.1	3.5
3.3	3.4
3.2	3.9

The experiment result is shown in Fig. 1, where cluster number is five and iteration count is 10.

The experiment result in Fig. 1 shows that the same position belongs to the same cluster. All research interests are not forced to fully belong to a cluster. Similar graduate students belong to the same cluster. Similar teaching activities belong to the same cluster. Positions, research interests, and graduate students are not forced to fully belong to any one of the user profiles. Thus, the algorithm might perform well.

4.2. Comparative results and analysis

We study the performance of the FCOU method against the existing Trajkova and Gauch's method [16]. We adopt the $F_{0.5}$ measure to evaluate the performance of the FCOU method. The $F_{0.5}$ measure is a standard IR evaluation measure. The $F_{0.5}$ measure can weight precision twice as much as recall, that is, $F_{0.5} = (1 + 0.5) \cdot (\text{precision} \cdot \text{recall})/(0.5 \cdot \text{precision} + \text{recall})$. Recall, precision, and the $F_{0.5}$ measure are calculated for each user considering the top different percentages of all non-zero concepts ordered by number of Web pages in the user profile. The user profile was built by classifying the first 30 pages visited by a user. Figs. 2–4 show the precision, recall and the $F_{0.5}$ measure calculated for each user considering 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100% of the top ranked concepts. Figs. 2–4 show that the precision, recall and the $F_{0.5}$ measure of FCOU method is improved more obviously than that of the Trajkova and Gauch's method [16]. It is because

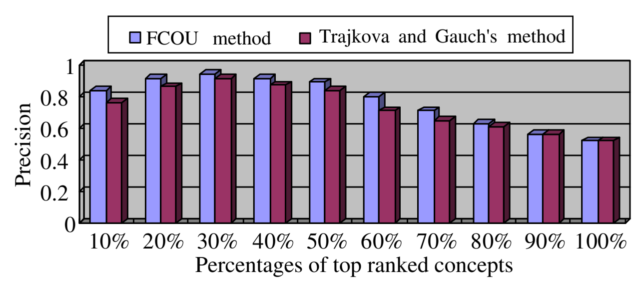

Fig. 2. The comparative result of precision.

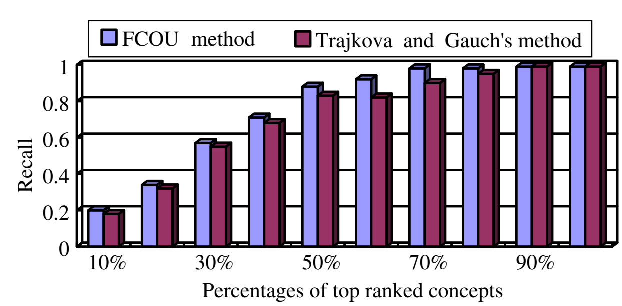

Fig. 3. The comparative result of recall.

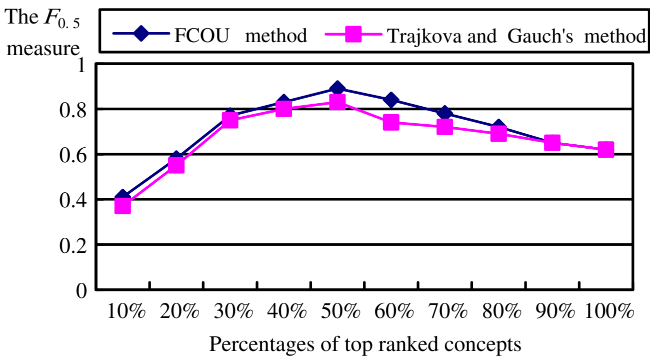

Fig. 4. The comparative result of the $F_{0.5}$ measure.

some information is allowed to belong to several user profiles simultaneously with different degrees of accuracy.

5. Conclusion

With the emergence of the Semantic Web, ontology offers some opportunities for improving on information retrieval system. Ontology-based user profiles typically maintain sophisticated representations of personal interest profiles. These representations can be utilized for effective information retrieval. In this paper, we propose a fuzzy clustering method, FCOU, of construction of ontology-based user profiles. The FCOU method employs the combination of FCM, PHR and simulated annealing to develop user profiles. In FCOU, an augmented Lagrangian function is used to create fuzzy clustering model for the construction of user profiles. The OSGP algorithm is used to compute the optimal solution under given a vector of Lagrange multipliers estimates and a penalty parameter conditions. The CPO algorithm is used to compute the optimal solution. The CCF algorithm is used to construct user profiles. Thus, the method can improve precision and recall.

Acknowledgements

The authors thank Professor Linping Sun for technical help. This work is supported by the National Natural Science Foundation of China under Grants 60673186, 60571048, 60873264 and 60736015, and the national 863 program under Grants 2007AA01Z178. The work is partly supported by China NSF grants (60573131, 60673154, 60721002, 60825205), Jiangsu High-Tech Research Project of China (BG2007039), and China 973 Project (2006CB303000).

References

[1] Kanjo D, Kawai Y, Tanaka K. A3: framework for user adaptation using XSLT. In: The proceedings of the thirteenth international world wide web conference (WWW2004). New York, USA, May 17–22; 2004. P. 252–3.
[2] Bezdek JC, Ehrlich R. FCM: the fuzzy *c*-means clustering algorithm. Comput Geosci 1984;10:191–203.
[3] Nascimento S, Mirkin B, Moura-Pires F. Multiple prototypes model for fuzzy clustering. In: Kok J, Hand D, Berthold M, editors. Advances in intelligent data analysis. Third international symposium (IDA'99), lecture notes in computer science, vol. 1642. Springer-Verlag; 1999. p. 269–79.
[4] Nascimento S, Mirkin B, Moura-Pires F. A fuzzy clustering model of data and fuzzy *c*-means. In: The ninth IEEE international conference on fuzzy systems: soft computing in the information age (FUZZ-IEEE 2000), IEEE neural networks council. San Antonio, Texas, USA, May 2000. p. 302–7.
[5] Tjhi W-C, Chen LH. Fuzzy co-clustering of web documents. In: Fourth international conference on cyberworlds (CW 2005), 23–25 November 2005, Singapore. p. 545–51.
[6] Ahn J-W, Brusilovsky P, Grady J, He DQ, Syn SY. Open user profiles for adaptive news systems: help or harm? In: The proceedings of the sixteenth international world wide web conference (WWW2007). Banff, Alberta, Canada, May 8–12; 2007.
[7] Hu J, Zeng HJ, Li H, Niu C, Chen Z. Demographic prediction based on user's browsing behavior. In: The proceedings of the sixteenth international world wide web conference (WWW2007). Banff, Alberta, Canada, May 8–12; 2007.
[8] Nébel I, Smith B, Paschke R. A user profiling component with the aid of user ontologies. In: Proceedings of workshop learning – teaching – knowledge – adaptivity. Karlsruhe; 2003.
[9] Razmerita L, Angehrn A, Maedche A. Ontology-based user modeling for knowledge management systems. In: Proceedings of the ninth international conference on user modeling. University of Pittsburgh, Johnstown, USA, June 22–26; 2003.
[10] Kim H-L, Kim H-G, Park K-M. Ontalk: ontology-based personal document management system. In: The proceedings of the thirteenth international world wide web conference (WWW2004). New York, USA, May 17–22; 2004.
[11] Nakaya N, Kurematsu M, Yamaguchi T. A domain ontology development environment using a MRD and text corpus, fifth joint conference on knowledge-based software engineering, frontiers in artificial intelligence and applications, vol. 80. IOS Press; 2002. p. 242–51.
[12] Maedche A, Staab S. Ontology learning for the semantic web. IEEE Intell Syst 2001;16(2):72–9.
[13] Li YF, Zhong N. Mining ontology for automatically acquiring web user information needs. IEEE Trans Knowl Data Eng 2006;18(4):554–68.
[14] Bouquet P, Serafini L, Zanobini S, Sceffer S. Bootstrapping semantics on the web: meaning elicitation from schemas. In: The proceedings of the fifteenth international world wide web conference (WWW2006). Edinburgh, Scotland, May 23–26; 2006.
[15] Suchanek FM, Kasneci G, Weikum G. YAGO: A core of semantic knowledge unifying wordnet and wikipedia. In: The proceedings of the sixteenth international world wide web conference (WWW2007). Banff, Alberta, Canada, May 8–12; 2007.
[16] Trajkova J, Gauch S. Improving ontology-based user profiles. RIAO 2004, Vaucluse, France, April 26–28. p. 380–9.
[17] Bezdek JC. Pattern recognition with fuzzy objective function. Plenum Press; 1981.
[18] Bezdek JC. Some non-standard clustering algorithms. In: Legendre P, Legendre L, editors. Developments in numerical ecology. NATO ASI series, vol. G14. Springer-Verlag; 1987.
[19] Zadeh LA. Fuzzy sets. Information and Control 1965;8:338–53.
[20] Metropolis Nicholas, Rosenbluth Arianna W, Rosenbluth Marshall N, Teller Augusta H, Teller Edward. Equation of state calculations by fast computing machines. J Chem Phys 1953;21:1087–91.
[21] Kirpatrick S, Gellat CD, Vecchi MP. Optimization by simulated annealing. Science 1983;4598:671–80.
[22] Cerny V. Thermodynamical approach to the traveling salesman problem: an efficient simulation algorithm. J Optimiz Theory Appl 1985;45(1):41–51.
[23] Gray P, Hart W, Painton L, Phillips C, Trahan M, Wagner J. A survey of global optimization methods. <http://www.cs.sandia.gov/opt/survey/main.html>. March 10; 1997.
[24] Jonhson DS, Aragon CA, Mcgeoch LA, Schevon C. Optimization by simulated annealing: an experimental evaluation – Part II (graph coloring and number partition). Oper Res 1991;39(3):378–406.
[25] Xie KX, Han LX, Lin YL. Optimization method. Tientsin: University Press; 1997.
[26] Wilkinson K, Sayers C, Kuno HA, Reynolds D. Efficient RDF storage and retrieval in Jena2. In: The first international workshop on semantic web and databases, co-located with VLDB 2003 (SWDB 2003). Humboldt-Universität, Berlin, Germany, 7–8 September 2003. p. 131–50.
[27] Carroll J, Dickinson I, Dollin C, Reynolds D, Seaborne A, Wilkinson K. The Jena semantic web platform: architecture and design, HP Laboratories Technical Report HPL-2003-146; 2003.