ELSEVIER

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Engineering 29 (2012) 780 – 784

Procedia Engineering

www.elsevier.com/locate/procedia

2012 International Workshop on Information and Electronics Engineering (IWIEE)

Study on Personalization Recommendation System Based on Recruitment Information

Tao Tan*, Leting Tan

School of Computer, China West Normal University, Nanchong 637002, China

Abstract

With the rapid development of Deep Web, a large number of Deep Web information often lead to "information overload" and "information disorientated ", yet, personalized service can solve this problem. This paper introduces personalization recommendation to the Deep Web data query, a user interest model based on fine-grained management of structured data and a similarity matching algorithm based on attribute eigenvector in allusion to personalization recommendation are proposed. Secondly, develops a prototype recommendation system based on recruitment information. Finally, the efficiency and effectiveness of the personalization recommendation are verified through the experiment.

Keywords: Personalization Recommendation; Similarity Matching; User Interest Model; Recruitment Information; Deep Web

1. Introduction

The data contained in web is significant in many fields. The information seems disorderly and unsystematic, but seeing from the depth of the web information, it can be divided into two parts: the Surface Web and the Deep Web. Surface Web refers to the Web pages that can be searched by traditional search engines (e.g. Google, Yahoo, etc) through hyperlinks. Deep Web refers to the content that can not be indexed by traditional search engines, and these contents mainly stored in the online web databases.

* Corresponding author. Tel.: +86-0817-2254525.
E-mail address: tantao99132@163.com.

doi:10.1016/j.proeng.2012.01.041

With the arrival of the era of web2.0, web database is widely applied in all kinds of dynamic web site, and web is also boosting its "deepening" process.

At present, the main way to obtain information from the Deep Web is to send a query through the query interface of dynamic web pages. Because of the special accessing method to web database, the content can not be indexed effectively by the traditional search engines. In order to help users effectively use tremendous information of the Deep Web, the researchers started researches on Deep Web data integration. Deep Web data integration can partly satisfy people's needs of information query in the Deep Web, but it is not intelligent, and cannot learn users' interest. As for the ones have specific professional interest, information needs to remain unchanged in quite a long period of time or changes a little, and users can only search the same content online repeatedly, causing much unnecessary waste. Driven by this demand, personalized service technology got rapid progress. If the query based on Deep Web data integration is in the service form of "people seeking information", then personalization recommendation is in the service form of "information seeking people". Find out users favor according to their behavior patterns, dynamically custom the query content for them and that will bring users great convenience.

This paper puts forward a user model based on fine-grained management of structured data. And build initial user interest model through the five element sets of user. Aimed at personalization recommendation, a similarity matching algorithm based on attribute eigenvector is proposed. On the basis of the ideas above, we developed a prototype recommendation system based on recruitment information and the experiments prove efficiency and effectiveness of the personalization recommendation .

2. Related Principles

Deep Web consists of four basic components: web site, database, query interface and hyperlink. A website is a server, maintaining one or many backend database that saving addressable online information. Each backend database can be queried through one or many HTML Form; these query forms build their own query interfaces [9]. Each query interface supports several attributes, for example, if you are searching for a position through a job query form offered by a recruitment information site, it can be searched according to these attributes such as job categories, industry categories and work sites. These fields, attributes together with some semantic tags constitute the schema information of query interface. When you make a choice or input a keyword and then submit, the web site will return the result pages including information that meet your queries to you.

Personalized service technology use user model to describe users' interest, calculate similarity of information and user model, recommend high correlation information to users. Because users' interests keep changing, so the construction of user interest model is a learning process, it updates according to the feedback information from users.

In order to get accurate and necessary information from large-scale dynamic Deep Web, scholars started researches on data integration in recent years. So far, they have made achievements in this field, such as query interfaces integration [1, 2], results data extraction and annotation [3, 4], etc. Moreover, personalization recommendation is getting mature in application, for instance, the construction of user interest model based on ontology[5],the user interest modeling and update strategy of extended semantic[6].They also made some achievements in the personalization recommendation based on Deep Web data. The literature [7] put forward a service frame based on the Deep Web personalized service, and it has been applied to the scientific literature retrieval, constructed the recommendation system of the science and technical literature.

Simultaneously, the development of Web enables network recruiting gradually becomes the main model for enterprise recruiting and individual job seeking. "Network and Mobile Data Management Laboratory" led by Professor Meng Xiaofeng in Information Institute of Renmin University of China use

recruitment information as an example to develop a "Job Tong" prototype system, At present, the "Job Tong" prototype system has been released on the internet, including recruitment information of 51 JOB, Chinahr.com, ZhiLian ZhaoPin, domestic universities and large work information websites. In the recruitment information retrieval service, as there are characteristics of two-way choice of talented person and the position between job seekers and employers, this paper designs the personalization recommendation into two-way recommended services. Both provide personalized job recommendations for job seekers, and provide personalized talents recommendation for recruiters, implementation ideas of the two way service are the same.

3. Constructing and updating of user interest model

3.1. Expression of User interest model

Users' interests are various and ever-changing, it would be too simply to describe them as interested or uninterested. It can neither effectively describe users' multiple interest features, nor timely tracking users' interests changing, especially some interests update frequently and change shortly. Considering the factors mentioned above, in the period from users submit information needs to log off, do integrated description about one's interests. Thereinto, that includes the process of the user interest model make dynamic update as demand adjusts, in order to achieve the purpose of reflecting user interest information needs timely and accurately. The paper constructs fine-grained management of structured data. The user interest model concludes and summarizes from bottom to up based on the domain ontology, the forming process is simple and the description is accurate. At the same time it can describe the users' interest differently, reduce the interference between different categories, help frequent short-term theme interest changes, and improve the precision of the model updating.

This paper describes the user interest model as five-tuple M: M={S, K, L, W, T}, thereinto, K=（K_u, K_c）; S stands for establishes state of the user interest model, K stands for the feature of the corresponding theme. It composed of two parts: the K_u stands for feature model before renewal; K_c stands for feature after dynamically updated. In the initial user model state S_0, there is no feedback updates, so $K_c=\varnothing$. L stands for K's semantics, described using the recruitment information. Such as position, industry, work experience, etc.; W means the feature weight of K; T means the updating time, which is primarily used to analysis user interests change. For example: M= {S_i, K_i, L_i, W_i, T_i}， S_i is the thematic state of user interest model on the dynamic adjustment renewal. The corresponding feature set K_i, $K_i = (K_{iu}, K_{ic})$. L= {L_{i1}, L_{i2}, ..., L_{im}}are the semantics of corresponding K_i. W_i, represents the feature weight of K_i, is a value between [0, 1]. T_i represents the update time of corresponding S_i.

3.2. Similarity matching of user interest pattern

After establishing the initial user interest model, similarity and personalization recommendation based on the matching results is what we need to work on. The basic idea is: regarded highly related information as seed, and extended neighbors in mode library using similarity, seek for similar interest information in the user interest library, so as to improve the recall rate. We also tried the statistics and machine learning method to improve the classic similarity so as to get better effect.

Definition1. The query Q submitted by users: a Deep Web query was composed of a group of key words, $Q= \{q_i|q_i\in Q, 1<=i<=k\}$, thereinto, Q is the key word set for the users' query.

Definition2. The web database query interface WDBI: it was composed of attribute name, data types and the corresponding candidate value. Definition as follow： $WDBI=\{<A_i,ADT_i,AVA_i>|A_i\in A,ADT_i\in ADT,AVA_i\in AVA,1<i<k\}$. Among them, A stands for the

attribute name set of query interface of the Web database; ADT stands for the data type of the corresponding attributes, including TEXT, NUMBER and DATETIME; AVA stands for the candidate value set of corresponding attributes.

Definition3. The given query set Q_1 and the state set S_2 in user interest model, Q_1 and S_2 contains separately m and n data. The weight of each data in Q_1 is respectively: $w_{q1}, w_{q2}, \ldots w_{qm}$, The weight of each data in S_2 is respectively: $w_{s1}, w_{s2}, \ldots, w_{sn}$, Match the similarity of data q_i in Q_1 and data s_j in S_2, definition of the similarity measurement is: $Sim(Q_1,S_2)= ((\max \sum_{i=1}^{m}\sum_{j=1}^{n} Wqi * Wsj)*a_{ij})*v_{Ai}$ .Among these, the purpose of introducing a_{ij} is to make sure w_{qi} and w_{sj} participate in similarity match for only once. When w_{qi} and w_{sj} matches, $a_{ij} = 1$; Otherwise $a_{ij} = 0$.

Thereinto, v_{Ai} stands for the similarity metric of different attributes caused by different data types in the web database query interface. For TEXT types: we use vector composed of keywords to express, and suppose W_i to be the feature vector of user interest model S_i based on the attribute A_i. W_j stands for the feature vector of data query based on attribute A_i, we use the cosine value of vectors between intersection angles, and formula is:

$$v_{Ai}= (\sum_{k=1}^{n} w_{ik} * w_{jk})/\sqrt{\sum_{k=1}^{n} w_{ik}^2 \sum_{k=1}^{n} w_{jk}^2} \quad (1)$$

Generally, when publishing recruitment information, publishers usually were asked to fill in the corresponding information as fields on the web in order to adopt the database means, and information would be displayed in sub area of web page. At this moment recruitment information query can get all the domain information through the analysis that custom-tailored templates made to the web page HTML or XML code [8]. As for this kind of interest theme recommendation, this paper classify users' record by recording users' browse and query, and count users' interest degree, so as to analyzes and get their high interest degree and finally to determine the user interest theme.

Personalization recommendation based on the Deep Web uses the user queries as training set. Through the similarity matching between query data and user interest model, recommend higher interest degree to the users.

3.3. Updating user interest model

User interest model updating strategy: When the system detects the user interest model status, it will query the theme characteristics under the status, if this model contains these features, it is enough to change its weight value and update time. If the interest is not included in the user model and user interest model has surplus storage space, it will generate a new interest model directly. If storage space is full, we need to eliminate the small weight of interest in the model state to add some new ones.Feature weight value updating strategy: Suppose D to be the Deep Web data set with additional feedback information, $d_j \in D$, n_j stands for the browsing number of d_j, t_j stands for the browsing time of d_j, W stands for the weight value of feature item K. Weight value updates according to the formula (2) to adjust the system. That is to say according to the time interval between the current system time T_n and the original update time T of feature item and the users' browsing behavior to determine the new weight. At the same time, modify the corresponding update time to T_n.

$$w(T_n)= \frac{\alpha}{\alpha + (T_n - T)} * \sum_{dj \in D} (d_j * t_j) * W_T \quad (2)$$

$$T=T_n \quad (3)$$

4. Experiment

In order to give an objectively evaluate to the personalization recommendation methods based on Deep Web proposed in the paper, a prototype is realized according to the theory and algorithm above, and verify a web database in a real. In this experiment, we use three Deep Web data sources to carry out Deep Web data crawl: ZhiLian ZhaoPin, ChinaHR, and 51 Job. As for user query q_i, the number of recruitment information item in the three data sources are 878,997 and 539. For 15 registered users, each user has five initial interests, record and return data start from the users' first query. For each interest, the maximum return information item number that system setting gets from Deep Web data resource is 30.

In this paper the recall and precision are also regard as standard in personalization recommendation. For 15 registered users, each user has five initial interests, constructs the learning of the same kinds of interests and learning of the different kinds of interests. Supposed the initial interests of 15 registered users are all including "computer", records the learning model of the 15 users with the same interest. The experimental data shows that learning from the same kinds of interests, with the continuing learning of the user interest model, the precision and recall of recommendation are rising steadily. At the same time, the interest of each user will change, records the learning model change of user "user I", and the experimental data shows that as for the interest change of same user, the interest theme shows the tendency of dispersion. This is because the recent interest theme is always used to match in the recommendation algorithm.

5. Conclusion

With the rapid development of Deep Web, a large number of Deep Web information often lead to "information disorientated ", yet, personalized service can solve this problem. this paper put forward a solution that apply personalization recommended service based on the Deep Web to recruitment services, and that enables the users to obtain better personalized service with few participation. But, some details in this operation still needs further improvement and discussion in the future: as for users' satisfaction and quality of personalization recommendation algorithm still needs more reasonable evaluation methods.

References

[1] Wu WS, Yu C, Doan AH, Meng WY. An interactive clustering-based approach to integrating source query interfaces on the deep Web. In: *the 24th ACM SIGMOD Int'l Conf. on Management of Data*, Paris: ACM Press; 2004, p. 95−106.

[2] He H, Meng WY, Yu C, Wu ZH. WISE-Integrator: An automatic integrator of Web search interfaces for e-commerce. In: *the 29th Int'l Conf. on Very Large Data Bases,* San Fransisco: Morgan Kaufmann Publishers; 2003, p.357−368.

[3] Zhai YH, Liu B. Web data extraction based on partial tree alignment. In: *Proc. of the 14th Int'l World Wide Web Conf*, Chiba: ACM Press; 2005, p. 76−85.

[4] Zhao HK, Meng WY, Wu ZH, Raghavan V, Yu C. Fully automatic wrapper generation for search engines. In: *the 14th Int'l* World *Wide Web Conf*, Chiba: ACM Press; 2005, p. 66−75.

[5] Li Xueqing,Liu Ruihua. An approach for User Interest Model Based on Ontology. *Information Research.*

[6] Li Shan. The Representation and Update for User Profile in Personalized Service. *Journal of the China society for scientific and technical information* 2010; 29: p.67-71.

[7] Xu Hexiang. Research on Deep Web Integration and Its Related Several Technologies. *Dr. Dissertation* 2008.

[8] Wang Jingfan. Two-Step Job Information Retrieval based on Document Similarity. 2007.

[9] Liu Wei, Meng Xiao-Feng and Meng Wei-Yi. A Survey of Deep Web Data Integration. *Chinese Journal of Computers* 2007; 30: p.1475-1489.