A Context-Awareness based Dynamic Personalized Hierarchical Ontology Modeling Approach

Qian Gao\textsuperscript{a, *}, Xiangjun Dong\textsuperscript{a}

\textsuperscript{a}School of Information, Qilu University of Technology, \#3501 Daxue Road, Changqing District, Jinan 250353, P.R.China

Abstract

This paper proposes a new approach to construct an efficient and dynamic updated ontology model. In order to capture the implicit context information and to implement the personalized scalability demands of context model for different users, this paper proposes a new approach to construct a context-awareness based dynamic personalized hierarchical ontology model which divides the ontology into 2 layers: the first layer (general context-aware ontology) is used to capture the most essential conceptual entities in the context-aware computing environment, and the second layer (personalized ontology) is used to capture the individual user’s preference by considering the domain ontology, knowledge structure, the term co-occurrence frequency and the Active Degree of each device. An algorithm of Domain Classification Algorithm is used and an Improved Concept Identification Algorithm is proposed in this research. The simulation results show that the time overhead of the hierarchical ontology model is low and it provides the users with results that more accurately satisfy their specific goal and intent.

Keywords: Ontology, Context-Awareness, Preference, Personalized

1. Introduction

The future of the development of today’s Internet is gradually converted from a "computer-centric" to "user-centered". It is urgent to provide users with more accurate, efficient and personalized cross-media information processing and retrieval service. To solve this problem, a user preference profile approach has been utilized to offer each user personalized search results\textsuperscript{1}. Due to the advantage of unambiguous semantic expression,
knowledge sharing, reuse, interoperability, and a variety of effective reasoning mechanisms, the ontology is used by many researches to model data. However, the general domain ontology is not effective to capture the implicit context information and hence fails to implement the personalized scalability demands of context model for different users. In order to solve this problem, a personalized context-based hierarchical ontology model construction method is adopted to reconstruct the ontology with weighted related terms, so as to reflect the users’ preference. It is divided into 2 layers; the first layer called general context-aware ontology is used to capture the most essential conceptual entities in the context-aware computing environment, and the second layer called personalized ontology is used to capture the individual user’s preference by considering the domain ontology, knowledge structure, and the document information stored on all of the smart devices that belong to the same user, so as to realize the dynamic update of the general context-aware ontology.

2. The Realization of the Proposed Method

2.1 General context-aware ontology construction model

2.1.1 Time context-aware ontology model

Context can be divided into instantaneous context and continuous context based on time sequence features. Instantaneous context requires the description of point-in-time, while continuous context requires the portrayal of time interval. In this paper, we adopt the method proposed by Xu who established the time context-awareness ontology via the description mechanism combined with point-in-time and interval, and obtained the corresponding time information by realizing the connection with time ontology by means of the object properties of "has Time".

2.1.2 Location context-aware ontology model

The location information includes the position and the spatial topological relation of the users or the devices. In this paper, we integrated and simplified the SOUPA-Location ontology and the GCACO-Location ontology proposed by Chen and Xu, classifying the location ontology information into two kinds: one is Geographic Position class, and the other one is Space Region class. The Geographic Position class corresponds to the space coordinate system, and it supports the transformation and coordination of a variety of spatial reference systems. The Space Region class is used to describe the geometric features of a certain space area. One Space Region includes a series of Geographic Positions. Furthermore, the Space Region is further divided into office places and leisure places. The office location information is associated with device ontology, which is helpful for knowledge reuse. The location context-aware ontology model is shown in fig.1.

2.1.3 User context-aware ontology model

“Users”, the core concept in the computer environment perceived by the context, can exert certain influences on the tasks of administrators. This paper simplifies the ontology of GCACO users proposed by Xu, covering only the basic features and schedule of users. “Users” here refers to the behavioral entities in the computer context perceived by the context, including human and artificial intelligence agents. Within the model, the basic features of the user, including the personal data (such as name, status and age) and social nature (such as the acquaintances, relatives and colleagues), is described by “user profile”. This kind of information does not change frequently. As regard applications excluded from the special users, all required is to generally set the “user profile” target to describe the basic information of each user, rather than getting to know their specific features. The schedule of users is a collection of events, describing the time, place, participants and what happened, with each event involving one or more tasks. When a user is performing a specific task, this task may be decomposed into one or more activities. The above concepts need to correlate with the time and position entities to obtain relevant time and position information. Fig.2 illustrates the main concepts and relations in user context-aware ontology model.
2.2 Personalized context-aware ontology construction agent

The second layer is extended with co-occurrence terms and semantically related terms that are added as vectors for each concept of the ontology. The terms are given weights to reflect the importance of the semantic relation between the concept and the terms.

2.2.1 Definition of Personalized Ontology Profile
A personalized ontology profile is a Concept Vector with each concept in the ontology with weighted related terms. Suppose \( C \) is the concept set of the corresponding ontology. \( T \) is the set of \( n \) terms in the document collection used for the construction of the personalized ontological profile. \( t_i \in T \) denotes related weighted term \( i \) in the set of terms. Then the ontology profile for concept \( j \) is defined as the vector \( C_j = [w_1, w_2, \ldots, w_n] \) where each \( w_i \) denotes the semantically related weight for each term \( t_i \) with respect to concept \( C_j \).

### 2.2. 2 Domain Classification Algorithm

Before analysing the document stored in the user device, Performing Part of Speech (POS) \(^4\), Stop of Words Removal, and Words Stemming are used to to pre-process the text, and then the Domain Classification Algorithm (DCA)\(^5\) is used to identify which domain the document stored in the user device belongs to. The DCA is based on WordNet domains by the comparison between the domain of the document context and the domain of the word’s sense.

### 2.2. 3 Concept Identification Algorithm

An Improved Concept Identification Algorithm (ICIA) is further used to search the weighted corresponding concept in the WordNet to identify the concept in the general domain obtained in the first layer that corresponds with the one with the words in the document. The pseudo code of algorithm ICIA is shown in Algorithm ICIA.

**Algorithm ICIA**

Input: all of the words in document D gotten from Words Stemming phase.

Output: Set of General Domain concepts belonging to terms (words) in documents D.

Procedure: // CTi is the context of the words in the documents, it is the sentence in documents D that contains the word occurrence being analyzed.

1. Do   Get General Domain entries Ci set (C1, C2, C3, …) that contains the word Wi.
2. Store Wi in the corresponding Ci in the database.
3. While i<=N
4. Rank the concepts Ci according to the length of the concept in descending order.
5. For each Ci
6. {  Get common words between CTi and representative terms of Ci, which is the intersection CT=\( \cap \) (CTi, Ci)
7. If |CT|<|Ci| then   The concept sense Ci is not within the context CTi.
8. If |CT|=|Ci| then
9. { The concept sense Ci is within the context CTi.
10. Add Ci to the set of possible senses associated with the document }

### 2.2.2.3 Local Document based construction

In order to evaluate the importance of each term in the term set, so as to reflect the user’s daily behaviour, we weigh each added related term of the concept according to the term co-occurrence frequency in the corresponding general domain ontology.

In this stage, the personalized context-aware ontology model is constructed based on a document collection covering the same domain as the general ontology. We use the frequency of the terms found in the documents assigned to each concept, and calculate the co-occurrence weight of each term of the corresponding concept by the TF -based Score, as shown in Eq. 1:

\[
\text{cscore}_{i,k} = \alpha \frac{\sum_{j \in D} T_{f_{j,k}} \cdot \log(1 + \max(T_{f_{i,k}}))) \cdot \log \frac{N}{n_i}}
\]

Where \( T_{f_{j,k}} \) is calculated by Eq.2:

\[
T_{f_{j,k}} = \beta \cdot \sum_{d \in D} T_{f_{d,j}} + \lambda \cdot \sum_{p \in P} T_{f_{p,j}} + \delta \cdot \sum_{i \in S} T_{f_{i,j}}
\]
Tfi,d, Tfi,p and Tfi,s is the term frequency for term i in document, paragraph p and sentence s respectively. Tfk,i is the term frequency for term i of concept k. β, γ, δ is the constant which satisfies β < γ < δ and β + γ + δ = 1. N is the number of concept, ni is the number of the concept vectors containing term i, max(Tfh,k) is the frequency of the most frequent for the term h in the concept vector k. α is a constant modifier used to make the value of cscore < 1. For each extension term i for a certain concept k, the cscore weight smaller than 1 is discarded.

2.2.2.4 Device Activity based construction

The longer the intelligent device is online, the bigger is its contribution in determining the individual preferences of the user, and the bigger is the corresponding weight of the illustration term set generated by the internal storage of its document and browsing histories. The reformulated weight of the extension term set is created by Eq. 3:

\[
DW_{jk} = \sum_{t \text{device alive}} \frac{UserID.AD_{j}}{\sum_{t \text{all device of UserID}} UserID.AD} \cdot W_{ijk,t}
\]

Here, UserID.AD_j is the Active Degree of device j that belongs to UserID. W_{ijk,t} is the weight of the term k in concept C_{ij} which was calculated by device t. The weight smaller than the threshold 0.7 is discarded.

3. Simulation

3.1 Experimental Data Sets

The used ontology data include the open standard ontology CYC Upper Ontology and the hierarchical context-awareness ontology proposed in this paper. Office domain ontology and leisure domain ontology were constructed in the simulation. All the domain ontology was modelled by the open and free graphical ontology modelling tool Protégé by OWL language. Furthermore, we use a branching factor of Open Directory which contains 160 concepts in the hierarchy and a total of 1789 documents indexed under various concepts.

3.2 Experimental results and analysis

The construction overhead of the hierarchy context-awareness ontology is shown in fig.3. From the figure we can see that the loading time of the upper general ontology is small, while that of the under layer leisure ontology increases with the growth of the number of the concepts and the number of the properties. But the merge time of the two layers are both low, which means that the hierarchy structure can control the size of the ontology effectively, so as to reduce the burden of the system.

The performance of the hierarchical ontology model based context reasoning is shown in fig.4. From the figure we can see that, the reasoning execution time of the hierarchical ontology model increased along with the growth of
the size of the ontology, but the growth rate is lower than the non-hierarchical ontology; and for the same size, the reasoning execution time of the hierarchical one is less than the non-hierarchical one, because the general upper-layered ontology with larger scale only need load, verify and analyze once, and then the class and instances relevant to the domain ontology with smaller scale are needed to load, verify and analyze. Therefore, the hierarchical ontology can ameliorate the burden of the reasoning.

The performance of the experimental models was measured by the precision averages at 11 Standard Recall Levels (11SPR), and the result of the accuracy of the weight of each concept as well as the influence of the accuracy of the context reasoning is shown in fig.5.

It can be seen from the simulation result that since the proposed method comprehensively considers the general ontology and the Knowledge Structure as well as the users' behavior, the weights of the concepts in the personalized context-aware ontology model are updated all the time and are adapted to changes in user interest rapidly. Therefore, the proposed method provides the user with results that more accurately satisfy their specific goal and intent for the search.

4. Conclusion

This paper proposed a new approach to construct a context-awareness based dynamic personalized hierarchical ontology model which divides the ontology into 2 layers: the first layer is the general context-aware ontology, and the second layer is the personalized ontology which is extended with the co-occurrence terms and semantically related terms that are added as vectors for each concept of the ontology. Simulation shows that the proposed hierarchy structure can control the size of the ontology effectively and reduce the scale of the loading and analysis of the upper general context-aware ontology, so as to reduce the burden of the context reasoning.

Acknowledgements

This work was supported partly by National Youth Science Foundation of China (61502260), National Natural Science Foundation of China (71271125) and Natural Science Foundation of Shandong Province, China (ZR2011FM028).

Reference

2. Xu nan, Research on Key Techniques of Context Aware Computing based on Ontology, Dissertation, Dalian Maritime University, 2015