MICISO: Towards High Performance Ontology Learning

Liu Yang [1], Guojie Li [1], Zhongzhi Shi [1]
[1] Institute of Computing Technology Chinese Academy of Sciences, 100080 Beijing, China
yangliu@ustc.edu, lig@ict.ac.cn, shizz@ics.ict.ac.cn
http://www.ict.ac.cn/

Abstract

We put forward a hypothesis that there exist common meaningful structures among ontologies whose domains are analogous to each other. The initial motivation of our hypothesis is to make full use of the structural information in existing ontologies, in order to guide the process of ontology learning and improve its quality especially in structural aspect. To support the hypothesis we give a series of definitions, algorithms and some preliminary experiments. We suppose that our work will spark a novel promising thinking for the domain of ontology, i.e. to study existing ontologies for useful things.

1 Introduction

There are a number of methods for such ontology learning today [Maedche, 2002], however, the quality of these methods is not so satisfactory [Gomez & Manzano, 2003]. We infer that an important reason is that they have not made full use of the existing ontologies. With this idea we put forward a hypothesis that there exist some common meaningful structures among ontologies whose domains are analogous to each other. If such structures are obtained, we can apply them as guidelines in ontology learning, in order to improve the structural quality of the results. We would like to lead people to the scene during construction of ontologies demonstrated in fig.1.

2 Definitions

We suppose that such common meaningful structures exist in different ontologies, however, they are equal from the point view of graph theory. Meanwhile, we prefer such structures as complex and large as possible. Based on the above consideration we present the definition as follows.

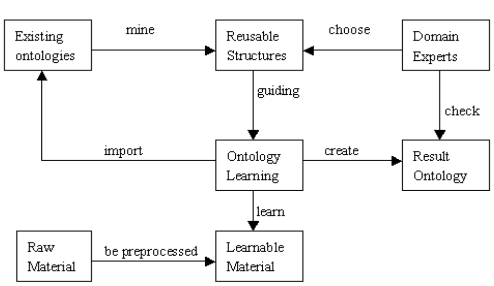

Fig. 1. The proposed model for ontology learning with our hypothesis

We will begin with the definition of ontology with reference of the description of ontology [Gruber, 1993], the semantics of OWL [Patel *et al.*, 2004].

Definition 1Core Ontology (CO) A core ontology is a structure $O_C=(C, \leqslant_C, R, \sigma_R, P, A)$, which omit instances, data types, and values in ontology.

Definition 2Induced Sub-Ontology For any ontology $O_0=(C_0, \leqslant_{C0}, R_0, \sigma_{R0}, P_0, A_0)$, $O=(C, \leqslant_C, R, \sigma_R, P, A)$, if C_x is a concept set and P_x is a predicate set, O is called the induced sub-ontology of O_0 on concept set C_x and predicate set P_x, which is denoted as $O=\pi(O_0 \mid C_x, P_x)$, iff (1)$C=C_x \cap C_0$ (2)$\leqslant_C=\leqslant_{C0}\cap C\times C$ (3)$R=\{ r \mid r\in R_0$,and for any $1\leqslant i \leqslant |\sigma_{R0}(r)|, \pi_i(\sigma_{R0}(r)) \in C \}$ (4) σ_{R1} is the restriction of σ_{R2} on R (5)$P=P_x\cap P_0$ (6)$A=\{ a \mid a\in A_0$,the predicate of A is from P,and the individuals of A are all from $C\cup R \}$

Definition 3Ontology Isomorph For ontologies $O_1=(C_1, \leqslant_{C1}, R_1, \sigma_{R1}, P_1, A_1)$, and $O_2=(C_2, \leqslant_{C2}, R_2, \sigma_{R2}, P_2, A_2)$, we call that O_1 is isomorphic with O_2 , denoted as $O_2 \approx O_1$, iff there exist two bijuctions f: $C_1 \rightarrow C_2$,g: $R_1 \rightarrow R_2$ and a mapping h:$R_1 \rightarrow N$, where $N=\{ k \mid k\in K[n]$,and $n\in Z^+$, K[n] is the set of all the n-rank permutations }, so that (1)For any $c_1,c_2\in C_1$,$c_1 \leqslant_{C1} c_2$ iff $f(c_1) \leqslant_{C2} f(c_2)$ (2)For any $r\in R_1$,$|\sigma_{R1}(r)| = |\sigma_{R2}(g(r))|$, and for any $1\leqslant x\leqslant |\sigma_{R1}(r)|$,if $y = [h(r)](x)$, then $\pi_x(\sigma_{R1}(r))= f(\pi_y(\sigma_{R2}(g(r))))$ (3)$P_1 = P_2$ (4)$a\in A_1$ iff $a|(f,g,h) \in A_2$, denoted as $O_1|(f, g, h)= O_2$.

Definition 4 Isomorphic Common Induced Sub-Ontology (ICISO) For an ontology set $S=\{O_i|O_i=(C_i, \leqslant_{Ci}, R_i, \sigma_{Ri}, P_i, A_i),1\leqslant i\leqslant n \}$, a predicate set P, and an ontology O, O is called a common isomorphic sub-ontology of S on P, iff for any $1\leqslant i\leqslant n$, there exists a concept set C_i', which is a subset of C_i ,so that $\pi(O_i \mid C_i', P) \approx O$. And the set of all the ICISOs of S on P is denoted as Com(S, P).

Definition 5 Maximum Isomorphic Common Induced Sub-Ontology (MICISO) For an ontology set S and a predicate set P, an ontology O is called a maximum common isomorphic sub-ontology, iff (1) $O\in$ Com(S,P) (2)There exists no O' so that $O'\in$ Com(S,P), $O'\neq O$,$O \leqslant_O O'$. The set of all the MICISO of S on P is denoted as MC(S, P).

3 Search MICISO

Based on the definition of MICISO, we are to specify the problem of search MICISO, and further design algorithms for it.

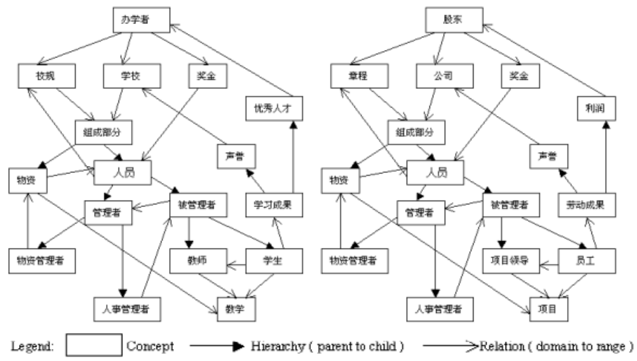

Fig. 2. Sketch map of the result

We will try to find MICISOs between only two original ontologies in consideration that firstly analogy usually exists between two domains, and even for more-than-two cases, we can use stepwise method in order to apply pruning and filtering strategy as soon as possible.

The problem is to derive all the MICISOs of the two ontologies on the predicate set, when two ontologies and a predicate set are given. In fact, it is a novel data-mining problem, thus we can it MISICO mining

As a data-mining problem, the algorithm can be extended from Apriori algorithm [Agrawal & Srikant, 1994]. Apriori algorithm derives all frequent itemsets in ascending order of the size of the itemset based on the monotonic property that the support of an itemset is less or equal to the supports of its subsets. Here, similarly for any induced sub-ontology O' of O, if O∈Com(S,P) we have O'∈Com(S,P). With this property, ICISOs are derived stepwise in ascending order of their sizes beginning with the simplest ICISOs, each of which has only one concept, and those ICISOs that cannot be expanded constitute the results, as is described above. Firstly, we simplify the two ontologies respectively by projecting them on the predicate set P. Secondly, the ICISOs with only one concept are all created, and they build the set C_1. Next, different pairs of ICISOs in C_n are tested for combining them to new ICISOs for C_{n+1}. Thus, C_{n+1} is generated from set C_n; and some of the members in C_n, which cannot be extended by combination with any other one and are possibly meaningful, form F_n. Therefore the members of F_n are MICISOs with n concepts. Above process are repeated until C_n becomes empty. Finally the set $\cup_n F_n$ are returned as the result.

TABLE 1

CORRESPONDING CONCEPTS AND RELATIONS IN THE RESULT

Ontology on University			Ontology on Corporation		
Relation	Domain	Range	Relation	Domain	Range
Stipulate	Founder	School-regulation	Stipulate	Shareholder	Constitution
Own	Founder	University	Own	Shareholder	Corporation
Donate	Founder	Bonus	Provide	Shareholder	Bonus
Administrate	School-regulation	Component	Control	Constitution	Component
Encourage	Bonus	Personnel	Encourage	Bonus	Personnel
Include	University	Component	Include	Corporation	Component
Co-stipulate	Personnel	School-regulation	Co-stipulate	Personnel	Constitution
Promote	Staff	Manager	Promote	Staff	Manager
Provide	Material	Personnel	Provide	Material	Personnel
Manage	HR-Manager	Staff	Manage	HR-Manager	Staff
Manage	Material-Manager	Material	Manage	Material-Manager	Material
Grow	Student	Tutor	Develop	Stuff	Project-leader
Provide	Tutor	Education	Lead	Project-leader	Project
Learn	Student	Education	Engage in	Stuff	Project
Ensure	Material	Education	Support	Material	Project
Generate	Student	Learning result	Generate	Stuff	Result
Enhance	Reputation	University	Develop	Fame	Corporation
Reward	Talents	Founder	Reward	Margin	Shareholder

Furthermore we create a random version of the algorithm in order to reduce the computing complexity, which will not be presented for lack of space.

4 Preliminary Experiment

Based on the algorithm above, we have developed a practical tool for MICISO mining and result checking. With our tool, the procedure of MICISO mining two original ontologies is as follows. Firstly, build a new project and add the two ontologies to the project; secondly, customize a mining task, i.e. appoint two original ontologies, select a set of predicates, choose the maximum number of result ontologies, and set up some heuristic rules for filtering results; thirdly, launch the task, and several result ontologies will be recommended; finally, you may check the result by viewing the graphs of their structures, and compare them with the original ontologies.

We have implemented MICISO mining on several pairs of existing ontologies, and get some interesting result. Specifically, we will take a pair of ontologies, whose domains are corporation and university respectively, for example. We demonstrate the MICISO by displaying its corresponding sub-ontologies in the two original ontologies respectively in fig.2, and the detailed description in English is listed in table 1.

5 Conclusions

In this paper, we put forward a novel hypothesis in order to benefit ontology learning with MICISO. To support it, we give a precise definition of MICISO, design algorithms for searching MICISOs. Based on them we have developed a practical tool for mining and checking such structures, with which we take preliminary experiments on several pairs of existing ontologies and obtain some interesting result.

Our future work is engaged in two aspects. (1) To implement our algorithm on more ontologies to further verify our hypothesis, while to gain more MICISOs for future usage. (2) To implement further study on the structural information of the existing ontologies to strengthen the hypothesis.

References

[Gruber, 1993] Gruber, T. R. 1993. *A translation approach to portable ontology specifications*. Knowledge Acquisition 5(2):199--220.

[Maedche, 2002] A. Midche and S. *Staab: Ontology Learning for the Semantic Web*, IEEE Intelligent Systems, 16(2), March/April 2001.

[Gomez & Manzano, 2003] Gomez-Perez and Manzano-Macho. *A survey of ontology learning methods and techniques*, 2003. http://ontoweb.aifb.uni-karlsruhe.de/Members/ruben/Deliverable%201.5.

[Agrawal & Srikant, 1994] R. Agrawal and R. Srikant. *Fast algorithm for mining association rules*. In VLDB, pages 487--499, Sept 1994.

[Patel *et al.*, 2004] Peter F. Patel-Schneider, Patrick Hayes, Ian Horrocks. *OWL Web Ontology Language Semantics and Abstract Syntax*. W3C Recommendation 10 February 2004. Available at http://www.w3.org/TR/2004/REC-owl-semantics-20040210.