A Knowledge Fusion Framework in the Grid Environment

Jin Gou¹, Jiangang Yang¹, and Hengnian Qi²

¹ College of Computer Science of Zhejiang University,
310027 Hangzhou, China
{goujin, yangjg}@zju.edu.cn

² School of Information Engineering of Zhejiang Forestry College,
311300 Hangzhou, China
qhn@zjfc.edu.cn

Abstract. The paper presents a knowledge fusion architecture based on the grid platform. The proposed framework suggests a semi-structural paradigm that emphasizes connotation of distributed knowledge resources in a grid environment. Our approach involves an extractive process of meta-knowledge sets which predigests the diversion among multi-source knowledge, the Genetic Fusion Algorithm which can generate a new knowledge space, a resources allocation method with meta-knowledge directory service. Experimental results of a case study show the feasibility of design rationale underling knowledge grid.

1 Introduction

Knowledge fusion is an important component of the knowledge science and engineering, which can transform and integrate diversiform knowledge resources to generate new information [1]. So information and knowledge from distributed nodes can be shared and cooperating. Multi-agent technology and grid computation method are used to integrate specifically knowledge in a way, which require plenty of data conversion operations and mapping procedures [2][3][4]. The paper proposes the architecture based on meta-knowledge and ontology bases to replace the complex interchanging process among diverse knowledge bases with extraction of meta-information. In order to fuse knowledge according to not formats but connotation, the paper contributes the Genetic Fusion Algorithm (GFA).

Knowledge fusion will result in an enormous amount of knowledge resources on the web. In such settings, we encounter resources management and other challenges. Since grid technologies carry the promise to enable widespread sharing and coordinated use of networked resources with effective scheme, we adopt a kind of semi-structure data model to encapsulate data resources on the knowledge grid platform [5].

The purpose of this paper is to present a knowledge fusion architecture with an autonomous resource allocation method used in the grid and the GFA for generating a new knowledge space.

2 Architecture

The knowledge fusion is constructed on the Globus grid services. Figure 1 shows the overall architecture.

Fig. 1. Framework of the grid based knowledge fusion system

The Globus project is the infrastructure of Grid computation. Globus Toolkit[6] can run on several operating systems. When we extract meta-knowledge from any node, the ontology should be described underlying certain rules and added into the ontology base. In the following, we elaborate on this issue.

Ontology base is a complex description sets. It is much more difficult to model inference in such a distributed knowledge grid environment. Let O and O_i denote the ontology base and the i^{th} object instance in it. The O_i will be used in the paper, which is defined by

$$O_i = \{(P_j^i, T_j^i, D_j^i)\}. \quad (1)$$

where P_j^i denotes the j^{th} attribute of the object O_i , T_j^i denotes the type of P_j^i , D_j^i denotes its value, expression or behavior. The domain for variant j is decided by idiographic scope knowledge ontology.

Let S_k denotes meta-knowledge sets of the k^{th} knowledge base node which can be defined by

$$S_k = \{(C_1^k, E_1^k), (C_2^k, E_2^k), \dots, (C_n^k, E_n^k)\}. \quad (2)$$

where C_i^k denotes the i^{th} character object of S_k , E_i^k denotes the description content of C_i^k .

Relationships among character objects are not defined here because those will be described in the ontology base. In order to implement the interchanged process between meta-knowledge sets and knowledge space, the definiendum (C_i^k, E_i^k) stands for not embodied characters but denotative objects of knowledge ontology. That operation must be synchronous with the initialization of meta-knowledge sets. Meta-knowledge sets and ontology base are formalized to be fuse and generate new knowledge elements.

3 Fusion Algorithm

With a resource allocation method mentioned above, we contribute the GFA to generate a new knowledge space according to the embodiment of knowledge ontology.

Describe the supposition as constructing meta-knowledge sets, function as fusing and generating new knowledge space.

$$GFA(Fitness, Fitness_threshold, p, r, m) \quad (3)$$

where *Fitness* denotes assessing function for fusion matching grade, which can endow the given supposition sets with a matching grade. *Fitness_threshold* denotes the threshold value beyond which fusion process can not persist. p is a supposed number which should be adjusted according to the result of unifying among diversiform meta-knowledge sets. If dimension of some sets are too small, NULL can be used to instead of it. r and m are intercross and aberrance percent.

Some description of the algorithm can be found in [7], the major steps are summarized as follows:

Select: Choose any S_i and append it to H_s , let the counter $c=1$

if $c < (1-r)p$, do the following operation circularly:

choose any S_j not belonged to H_s from $\{S_k\} - S_i$

if $\exists (P_m^i, T_m^i, D_m^i) \in O_i$, and $P_m^i = "R(C_*^i, C_*^i)"$ and $D_m^i \neq \text{NULL}$

or $\exists (P_m^i, T_m^i, D_m^i) \in O_i$, and $P_m^i = "R(C_*^i, C_*^j)"$ and $D_m^i \neq \text{NULL}$

append S_j to H_s , $c++$.

Intercross: Selected result $\langle S_1, S_2 \rangle$ must make the following expression right:

$$E_1^i \cap E_1^j \neq \phi \quad (4)$$

this means the intersection of ontology relating to suppositions can not be temp. Any element except the first one meets the requirement above can be intercrossed.

Aberrance: For each (C_j^i, E_j^i) , reverse its value as follows:

for every member $S^k \neq S^i$ of H_s ,

if $O_i = \{P_m^i, T_m^i, D_m^i\}$ which can meet the requirement: $j \exists P_m^i = R(C_j^i, C_j^k)$

and $E_{j'}^k \neq E_j^i$

$E_{j'}^k$ is a result of reversing operation on E_j^i .

Solution knowledge can be generated as follows:

Create ontology corresponding with question (O_p) and meta-knowledge (restriction) sets (S_p). Search all knowledge states in K for S_a whose relationship grade to question state is the max. For each knowledge state related to question ontology, seek out its

relationship grade. It is also the percent of related knowledge points in knowledge state and those in question state.

Knowledge state S_k relates to question ontology O_p must meet requirement as follows:

$$\exists (P_j^p, T_j^p, D_j^p) \in O_p, (P_j^k, T_j^k, D_j^k) \in O_k.$$

Table 1. Result of a case study

	Transform times	Losing connotation	Generate new knowledge	Find solution	Can be reused	Uniform
Based on interchange	600	Yes	No	No	Can	Yes
Method in the paper	50	No	Yes	Yes	Can	Yes

4 A Referential Application and Summary

Let us shift focus beyond the abstract architecture and point to a case study.

As shown above, the framework in the paper can minimize cost in a knowledge grid especially when a knowledge fusion procedure runs on it. And it can also improve the reuse performance of knowledge elements.

We present a new knowledge fusion framework in a grid environment. Compared with traditional resource management system, method given in the paper gives more flexibility on task requirements and resource utilization. We have also presented the GFA which can fuse diversiform knowledge and generate new knowledge space according to the connotation of ontology. In the future, we will apply the framework to more finely granularity knowledge grid and optimize the matching process.

References

1. LU, R.Q.: Knowledge Science and Computation Science. Tsinghua University Press, Beijing China (2003)
2. James, M.: Structured Knowledge Source Integration and Its Applications to Information Fusion. Proceedings of The 5th International Conference on Information Fusion. Maryland: IEEE, (2002) 1340-1346
3. Mario, C., Domenico, T.: The Knowledge Grid. Communications of The ACM. 1 (2003) 89-93
4. Tomas, M., Zsolt, B., Ferenc, B., *et al*: Building an Information and Knowledge Fusion System. Proceedings of The 14th International Conference on Industrial and Engineering Applications of AI and Expert System. Budapest: ACM, (2001) 82-91
5. Ian Foster, Carl Kesselman, Steven Tuecke: The Anatomy of The Grid: Enabling Scalable Virtual Organizations. Lecture Notes in Computer Science, Vol. 2150, (2001) 1-26
6. The Globus Project <http://www.globus.org>
7. Mitchell, T.M.: Machine Learning. China machine press, Beijing China (2003)