Social Mediation for Collective Intelligence in a Large Multi-agent Communities: A Case Study of AnnotGrid

Jason J. Jung and Geun-Sik Jo

Department of Computer and Information Engineering
Inha University, Republic of Korea
jj2jung@intelligent.pe.kr, gsjo@inha.ac.kr

Abstract. Collective intelligence is a key issue for efficient collaborations on semantic grid environment. Through interactions among agents on the semantic grid, we want to discover social centralities underlying on agent communities. In this paper, we propose an ontology mapping approach to build collective intelligence based on social mediation. As a case study, we demonstrate *AnnotGrid* platform to support manual annotation tasks of image dataset.

1 Introduction

Collective intelligence is regarded as one of the important processes for collaborations between people, in order to solve a very complicated problem. They can share their own knowledge (or experiences) with each other, access the knowledge in the individual viewpoint, and integrate some relevant knowledge with theirs. Originally, this mechanism has been studied in the communities of social science and pedagogical science. In the context of computational and information science, we want to automatically build the collective intelligence among information systems (e.g., virtual organizations), and then, to provide meaningful services (e.g., recommendation) to people [1]. Particularly, the individuals are represented as not human actors but software agents, which are working on behalf of the corresponding persons [2]. The agents however have some drawbacks to communicate with others. Such problems are *i*) lack of computational power, *ii*) lack of information and knowledge, and even more seriously, *iii*) semantic heterogeneities.

For overcoming these problems, the agents can be reinforced with semantic grid computing [3]. The semantic grid is an extension of the current grid in which information and services are given well-defined meaning through machine-processable descriptions which maximize the potential for sharing and reuse. The existing grid environments suppose that the power of agents should be simply identical. We are motivated from the naive assumption. Depending on the ontological structure of agents, the semantic information provided during cooperations might be different from the agents.

In this paper, as shown in Fig. 1, we have investigated three-layered architecture for the semantic grid, which is composed of social, ontology, and resource layers. We propose a novel approach to integrate the intelligence reasoning capabilities based on

Fig. 1. Three-layered semantic grid architecture

social network analysis and co-occurrence analysis (i.e., extended *TF-IDF*) methods of the resources. The semantics delivered from an agent is applied to measure the strength of social ties with neighboring agents, because we have no idea about the relationship between two agents *A* and *B* on social layer.

The outline of this paper is as follows. In Sect. 2, we will define some notation and explain our ontology mapping algorithm for collecting the social ties among agents. Sect. 3 will propose community-based approach for social mediation on multi-agent environment. In Sect. 4, we demonstrate our *AnnotGrid* environment to support semantic collaboration. Finally, Sect. 6 give a conclusion and future work.

2 Constructing Social Network

Now we want to define agents and their community model. Social ties between the agents are discovered, so that they are aggregated to construct a weighted social network. The weighted link between two agents are computed from two steps; *i*) semantic similarity between ontologies that they are using, and *ii*) co-occurrence patterns of resources that they are possessing.

Definition 1 (Agent). An agent $\mathbb{A}_i$ in a social grid $\mathbb{SG}$ is composed of three modules; *i*) inference engine, *ii*) a set of ontologies $\mathcal{O}_i$, and *iii*) local resources $\mathcal{R}_i$ where each element $r_\alpha \in \mathcal{R}_i$ is annotated with $\mathcal{O}_i$.

In this paper, the inference engine is regarded as an annotation (or classification) system for local resources. Thereby, each agent can build his own “local” ontologies, and also he can exploit the “standard” ontologies such as OpenCyc [4] and SUMO¹. For convenience, $\mathcal{O}_i = \{\langle c_\alpha, rel, c_\beta \rangle | c_\alpha, c_\beta \in Annot(r_x), r_x \in \mathcal{R}_i\}$.

2.1 Semantic Similarity by Ontology Mapping

We can measure similarity between two agents $Sim^\mathcal{O}(\mathbb{A}_i, \mathbb{A}_j)$ by similarity measurement strategy on [5,6], which defines all possible similarities between classes Sim_C ,

¹ Suggested Upper Merged Ontology, <http://ontology.teknowledge.com>

relationships Sim_R , attributes Sim_A , and instances Sim_I . For simplicity, we need to use only Sim_C . Given a pair of classes from two different ontologies, the similarity measure Sim_C is assigned in $[0, 1]$. The similarity (Sim_C) between $c_\alpha \in \mathcal{O}_A$ and $c_\beta \in \mathcal{O}_B$ is defined as

$$Sim_C(c_\alpha, c_\beta) = \sum_{\mathcal{F} \in \mathcal{N}(C)} \pi_{\mathcal{F}}^C MSim_Y(\mathcal{F}(c_\alpha), \mathcal{F}(c_\beta)) \quad (1)$$

where $\mathcal{N}(C)$ is the set of all relationships in which the classes participates, and function $\mathcal{F}$ can measure label similarities (it can be replaced to string matching algorithms, e.g., edit distance). The weights $\pi_{\mathcal{F}}^C$ are normalized (i.e., $\sum_{\mathcal{F} \in \mathcal{N}(C)} \pi_{\mathcal{F}}^C = 1$).

In this paper, hence, we consider three relationships in $\mathcal{N}(C)$, which are the super-class (SUP), the subclass (SUB) and the sibling class (SIB), and Eq. 1 has to be extended to

$$\begin{aligned} Sim_C(c_\alpha, c_\beta) &= \pi_{\mathcal{L}}^C sim_L(\mathcal{F}(c_\alpha), \mathcal{F}(c_\beta)) \\ &+ \pi_{SUP}^C MSim_{SUP}(SUP(c_\alpha), SUP(c_\beta)) \\ &+ \pi_{SUB}^C MSim_{SUB}(SUB(c_\alpha), SUB(c_\beta)) \\ &+ \pi_{SIB}^C MSim_{SIB}(SIB(c_\alpha), SIB(c_\beta)). \end{aligned} \quad (2)$$

where the set functions $MSim_Y$ is formulated for comparing the two sets of entity collections. According to the characteristics of the sets, it is differently given by two equations

$$MSim_{SUP, SUB}(S, S') = \frac{\max_{\langle c, c' \rangle \in Pairing(S, S')} (Sim_C(c, c'))}{\max(|S|, |S'|)}, \quad (3)$$

$$MSim_{SIB}(S, S') = \frac{1 - \sum_{\langle c, c' \rangle \in Pairing(S, S')} (Sim_C(c, c'))}{\max(|S|, |S'|)} \quad (4)$$

where $Pairing$ is a simple matching function for generating all possible pairs from both sets. These equations express either positive or negative influences on semantic relationships between classes. It means that the more matched super- and subclasses can imply the corresponding classes are more similar, as shown in Eq. 3. In contrast, the sibling classes (Eq. 4) reflect the negative effect, because they make the semantics dispersed and the degree of similarity decreased.

2.2 Co-occurrence Patterns

Finally, we can construct a weighted social network as a set of agents by referring to the resource layer. In this paper, we assume that when agents annotate a resource with semantically similar classes, they are strongly correlated, i.e., the strength of social tie might be high. Thus, the strength of social tie between $\mathbb{A}_i$ and $\mathbb{A}_j$ is given by

$$w'_{e_{ij}} = \sum_{r_x \in \mathcal{R}_i \cap \mathcal{R}_j} \frac{\sum_{c_\alpha \in Annot_i(r_x), c_\beta \in Annot_j(r_x)} Sim_C(c_\alpha, c_\beta)}{Annot_i(r_x) \times Annot_j(r_x)}, \quad (5)$$

and the normalized strength is easily computed by $w_{e_{ij}} = \frac{w'_{e_{ij}}}{|\mathcal{R}_i \cap \mathcal{R}_j|}$.

Furthermore, there is another issue to enhance the weights between agents. This is an inactivation process by using conflicted annotation semantics. When Sim_C between annotations of a resource is very low, the strength of the social tie is supposed to be decreased. We want to discuss this issue for future work.

3 Community-Based Social Mediation

A set of agents which is participating to a semantic grid should be organized as agent communities.

Definition 2 (Agent community). An agent community $\mathbb{AC}_s$ is represented as $\mathbb{AC}_s = \langle \mathcal{N}, \mathcal{E} \rangle$ where $\mathcal{N}$ is a set of agents $\{\mathbb{A}_1, \mathbb{A}_2, \dots, \mathbb{A}_{|\mathcal{N}|}\}$, and $\mathcal{E}$ is a set of relations between the agents

$$\mathcal{E} \subseteq |\mathcal{N}| \times |\mathcal{N}|. \quad (6)$$

Each element is represented as $\langle e_{ij}, w_{e_{ij}} \rangle \in \mathcal{E}$ where weight $w_{e_{ij}} \in [0, 1]$ indicates the strength of a social link between two agents $\mathbb{A}_i$ and $\mathbb{A}_j$.

Property 1 (Degree of cohesion COH). A set of agents in an agent community $\mathbb{AC}_s$ have to be highly cohesive with each other. It is similar to subgroup identification. Based on [7,8], this constraint is simply formulated by

$$COH(\mathbb{AC}_s) = \frac{\sum_{\mathbb{A}_i, \mathbb{A}_j \in \mathbb{AC}_s} w_{e_{ij}}}{|\mathcal{N}|C_2} \geq \epsilon_{COH} \quad (7)$$

where ϵ_{COH} is a predefined threshold value.

Based on the social ties between two agents, we can apply a non-parametric approach, e.g., nearest neighborhood method [9]. As extending [10], this task is to maximize “semantic” modularity function $Q^\diamond$ on the social network of agents. Given the number of agent communities $\mathcal{S}$ from a set of agents, a social network of $\mathbb{SG}$ can be partitioned into a set of agent communities (or subgroup) $\{\mathbb{AC}_1, \mathbb{AC}_2, \dots, \mathbb{AC}_S\}$. The agents can be involved in more than one community. It means that an agent in $\mathbb{AC}_k$ can also be taken as one of members of $\mathbb{AC}_{k'}$. Based on Equ. 7, the modularity function $Q^\diamond$ is formulated by

$$Q^\diamond(\mathbb{SG}) = \frac{\sum_{i=1}^S COH(\mathbb{AC}_i)}{\mathcal{S}} \quad (8)$$

where all possible pairs of agents should be considered. Thus, the optimal formation of $\mathbb{SG}$ is discovered when $Q^\diamond(\mathcal{S})$ is maximized. For computing this, in this paper, we applied k -nearest neighborhood methods.

Within a certain $\mathbb{AG}_i$, the centrality of agents can be considered to find out which agent is more semantically powerful, i.e., which agents is the most relevant to make two heterogeneous agents be able to communicate and understand with each other.

Definition 3 (Centrality CTR). From a set of agents $\mathcal{AC}_x$, the centrality of an agent $\mathbb{A}_i$ is given by

$$CTR(\mathbb{A}_i) = \frac{\sum_{\mathbb{A}_j \in \mathcal{AC}_x} w_{e_{ij}}}{|\mathcal{N}| - 1} \quad (9)$$

where $\mathbb{A}_j$ can be in either the same agent community or not. Because the network of the agents are fully connected, it is based on closeness centrality [11] rather than others (e.g., betweenness and stress centralities).

Given the agents whose centralities is highest in the corresponding agent communities, We have to repeat the cohesion evaluation process. As repeating this centrality computing process, the $\mathbb{SG}$ is hierarchically built for efficient social mediation. For the social mediation based on semantic grid environment, we have to find out whether two agents are involved into a same agent community or not. The reason is because the centrality is available only in the same community. There are two cases, i.e., two red arrows meaning communications *i*) between heterogeneous agents A_1 and B_1 , *ii*) between A_1 and A_3 , as shown in Fig. 2. Within an agent group, the social mediation can be done

Fig. 2. Two kinds of social mediation based on centrality; Three agent communities are built, and agents C_1 , C_2 , and C_3 are assigned with the highest centrality in the agent communities, respectively

by the agent whose centrality is highest. For example, we can figure out that the communications between A_1 and B_1 in AgentCommunity₁ is possibly supported by C_1 . On the other hand, for communication between A_1 and A_3 , we need to know which collective intelligence should be integrated. In this case, agents C_1 , C_2 , and C_3 are possibly cooperating to integrate their semantic information to mediate between A_1 and A_3 .

4 Experimentation: A Case Study of AnnotGrid Environment

We have been implementing a semantic grid environment based on multi-agent systems, called *AnnotGrid*. Sharing a bunch of image files is usually regarded as a serious hurdle in distributed environment. In semantic web era, rather than content-based analysis, a variety of studies have been focusing on image annotation. The annotation is basically generated by human experts (or end-users). They have to put in some description about the images (somehow, they can refer to knowledge-based systems.).

The goal of this system is to support the annotation tasks to save their time and increase the precision of their descriptions. Basic scheme of *AnnotGrid* is as follows (For more details, see the project web pages²);

1. Building social grid with initial annotations; it is an initialization step of semantic grid.
2. Loading a new image r_i
3. Retrieving a set of existing annotation from other users; the agent can get the existing annotations about r_i from the other agents in both the same agent community and different ones.
 - local searching by himself
 - global searching by social mediator
4. Translating the retrieved annotation; The corresponding social mediators provide translation service to make the annotations understandable.

For evaluating this system, we invited 32 college students to annotate an image dataset³. Above all, because the class similarities built during annotating the images should become relatively large, we recommended to use some “‘standard’ ontologies such as

- Suggested Upper Merged Ontology (SUMO)
- WordNet (<http://wordnet.princeton.edu/>), and
- CIDOC CRM (CIDOC⁴ Conceptual Reference Mode, <http://cidoc.ics.forth.gr/>).

Each person had to initially annotate five images randomly selected by his own preferences, so that four communities had been built ($\epsilon_{COH} = 0.35$). As keeping annotating the imaged, the students were able to obtain annotations recommended by *AnnotGrid*, and feedback their opinions whether the recommendation is relevant or not.

Table 1. Experimental results on annotation recommendation by *AnnotGrid*

Index	User number	Number of annotation		Accuracy of annotation (%)	
		Local	Global	Local	Global
Community ₁	4	19	97	72.5	78.6
Community ₂	12	93	124	68.4	81.2
Community ₃	9	58	137	62.5	75.0
Community ₄	7	52	94	69.1	81.6
Average				68.1	79.1

As shown in Table 1, global searching facility has shown better performance than local one by approximately 11%. It proves our social mediations has been working properly. Especially, the users in Community₁ got the largest social mediation, even though the number of users are smallest. In case of large communities (2nd, 3rd, 4th communities), the users had acquired quite higher precision. We think that enough amount of semantic information in these communities make ontology matching process more precise.

² <http://intelligent.pe.kr/AnnotGrid/>

³ <http://intelligent.pe.kr/AnnotGrid/images/>

⁴ <http://www.willpowerinfo.myby.co.uk/cidoc/>

5 Related Work

There have been various annotation systems for not only images but also bookmarks and other multimedia [12,13,14]. Especially, many collaborative tagging systems, called Folksonomy, like del.icio.us and Bibsonomy (<http://www.bibsonomy.org>) have been trying to share the semantic tags with others. But, they do not have any facilities to analyze social features and automated recommendation process. Then, people have to directly inspect which tags (or semantics) can be exploited. In the aspect of computing architectures, SOA (service-oriented architecture) systems have been applied to data annotation [15]. Particularly, semantic overlay architecture for P2P network [16] is very similar to our scheme, because it can also hierarchical organize the linkage structure among peers.

6 Concluding Remarks and Future Work

In conclusion, we claim that the social relationships between agents should be discovered to support efficient cooperations on semantic grid environment. The proposed method are able to build hierarchical formation of people in a social network. For the practical experimentation, this paper demonstrates *AnnotGrid* system, which is capable of recommending semantic information for manual annotation task of photos.

As future work, we are planning to apply consensus method [17] to our system, in order to deal with conflicts between ontology mapping.

References

1. Jung, J.J.: Ontological framework based on contextual mediation for collaborative information retrieval. *Information Retrieval* **10**(1) (2007) 85–109
2. Maes, P.: Agents that reduce work and information overload. *Communication of the ACM* **37**(7) (1994) 30–40
3. Gil, Y.: On agents and grids: Creating the fabric for a new generation of distributed intelligent systems. *Journal of Web Semantics* **4**(2) (2006) 116–123
4. Lenat, D., Guha, R.: *Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project*. Addison-Wesley (1989)
5. Euzenat, J., Valtchev, P.: Similarity-based ontology alignment in OWL-Lite. In de Mántaras, R.L., Saitta, L., eds.: *Proceedings of the 16th European Conference on Artificial Intelligence (ECAI)*. (2004) 333–337
6. Jung, J.J., Euzenat, J.: Measuring semantic centrality based on building consensual ontology on social networks. In: *Proceedings of the Workshop on Semantic Network Analysis (co-located in ESWC 2006)*. (2006)
7. Wasserman, S., Faust, K.: *Social Network Analysis*. Cambridge University Press (1944)
8. Jung, J.J., Euzenat, J.: From personal ontologies to socialized semantic space. In: *Poster of the 3rd European Semantic Web Conference*. (2006)
9. Gowda, K.C., Krishna, G.: Agglomerative clustering using the concept of mutual nearest neighbourhood. *Pattern Recognition* **10**(2) (1978) 105–112
10. Newman, M.E.J.: Fast algorithm for detecting community structure in networks. *Physical Review E* **69** (2004) 066133

11. Sabidussi, G.: The centrality index of a graph. *Psychometrika* **31** (1966) 581–603
12. Kiryakov, A., Popov, B., Ognyanoff, D., Manov, D., Kirilov, A., Goranov, M.: Semantic annotation, indexing, and retrieval. In Fensel, D., Sycara, K.P., Mylopoulos, J., eds.: *International Semantic Web Conference*. Volume 2870 of *Lecture Notes in Computer Science*., Springer (2003) 484–499
13. Belhajjame, K., Embury, S.M., Paton, N.W., Stevens, R., Goble, C.A.: Automatic annotation of web services based on workflow definitions. In Cruz, I.F., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P., Uschold, M., Aroyo, L., eds.: *Proceedings of the 5th International Semantic Web Conference (ISWC 2006)*. Volume 4273 of *Lecture Notes in Computer Science*., Springer (2006) 116–129
14. Jung, J.J.: Exploiting semantic annotation to supporting user browsing on the web. *Knowledge-Based Systems* **xx(x)** (to appear) <http://dx.doi.org/10.1016/j.knosys.2006.08.003>
15. Aurnhammer, M., Hanappe, P., Steels, L.: Augmenting navigation for collaborative tagging with emergent semantics. In Cruz, I.F., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P., Uschold, M., Aroyo, L., eds.: *Proceedings of the 5th International Semantic Web Conference (ISWC 2006)*. Volume 4273 of *Lecture Notes in Computer Science*., Springer (2006) 58–71
16. Aberer, K., Cudré-Mauroux, P.: Semantic overlay networks. In Böhm, K., Jensen, C.S., Haas, L.M., Kersten, M.L., Larson, P.Å., Ooi, B.C., eds.: *Proceedings of the 31st International Conference on Very Large Data Bases (VLDB)*, ACM (2005) 1367
17. Nguyen, N.T.: Conflicts of ontologies - classification and consensus-based methods for resolving. In Gabrys, B., Howlett, R.J., Jain, L.C., eds.: *Proceedings of the 10th International Conference on Knowledge-Based Intelligent Information and Engineering Systems (KES 2006)*. Volume 4252 of *Lecture Notes in Computer Science*., Springer (2006) 267–274