Advanced Engineering Forum *Online: 2012-09-26*
ISSN: 2234-991X, Vols. 6-7, pp 501-505
doi:10.4028/www.scientific.net/AEF.6-7.501

Research of Ontology-based Agricultural Geographic Information Service Matchmaking

Bin Shi[1], Yeping Zhu[1], Feng Li[1] , Chunjiang Zhao[2], Yuchun Pan[2]

[1]Agricultural Information Institute of CAAS

Chinese Academy of Agricultural and Sciences

Beijing, China

[2]Beijing Research Center for Information Technology in Agriculture

Beijing Academy of Agriculture and Forestry Sciences

Beijing, China

E-mail: shib@nercita.org.cn, E-mail: zhaocj@nercita.org.cn

Keywords: OWL-S; Geographic Information Service; Semantic Service Matchmaking

Abstract. A semantic model is proposed in this paper for geographic information service composition. It descript the content of service and user request semantically based on ontology, it takes process model of composited service into account to improve the efficiency of service matching Algorithm. The matchmaker affords to obtain the services list as result to meet the intent of user by advanced matching Algorithm. The result of experiment shows that the method this paper proposed performs well.

Introduction

GIS since the 1960s, rapid development, continuously applied in the new domains and areas, but the sharing and exchange of geographic information has always been a tough problem faced by the GIS application system developers. In order to solve the problem, publish the GIS data with geographic information services (GI Service) based on the SOA architecture become the main means of technology. GI Service greatly enhanced openness, sharing capabilities and flexibility of geographic information, but it cannot achieve the semantic level of the share exchange, processing and aggregation. Semantic Web technologies can enhance the GI service semantic description capabilities, to improve the GI Service based on semantic information sharing capabilities. The method this paper proposed focus on the GI service semantic description of the W3C OWL-S ontology, the establishment of a semantic model of a unified GI Service, calculate the semantic similarity between the semantic description of services and user requests with WordNet ontology, get in line with user intent to improve the performance service matching Algorithm.

Related Work

Service matching is the matching process between the service request and service profiles. Web services have two kinds, atom services and the composite service. The composite service is made up by more than one atom may in the form of complex process, it's difficult to be expressed in the unified model. Researchers proposed different semantic matching method for the two kinds of services.

Atom Service Matching. Atom service is a functional unit that cannot be split again, so most of the method matches services with the input, output, precondition and effect (IOPE). Atom service matching methods have three types: (1) keyword based matching (2) concept similarity based matching, and (3) DL reasoning based matching. Keyword-based semantic expansion [1] with the dictionary ontology, such like WordNet, is helpful to improve the performance of matching algorithm. Paolucci [2] matches the services by their input and output, and classified the degree of

matching as four grades: *exact, plug in, subsumes and fail*, and this classification method is referenced by many researchers. But Paolucci only takes the IO into account. Luo [3] proposed a multi-level service matching algorithm, which tries considering the many factors to improve the performance of service matching. The algorithm calculates the similarity of service matching according to the classification of services, the IO of service, the precondition of service, the effect of service and QOS. Namgoong, H [4] also think that take the usability as indicator to measure the similarity of service matching on the basis of the IOPE. OWL-S and its predecessor DAML-S all are Description Logic (DL) based ontology, a lot of researches consider to take the advantage of the logic description language. It describes the service in DL [5], reasons and judges the matching degree between the two services. The disadvantage of this approach is the problem of logic unreachable.

Composited Service Matching. Besides the IOPE information, the semantic content of the composited services also includes the connection information of the list of atom services. Matching methods based on different composition methods can be classified as the following four types: (1) keywords / concepts based matching, (2) DL reasoning based matching, and (3) converting the process of composited service to graphics or other process description language and matching. Stroulia, E[1] considered a OWL-S composited services profile document as a whole, and represent document by VSM model to calculate the semantic similarity between the services and user request. But keywords / concepts-based approach ignores the importance of the composition structure of atom services. Using DL reason to match the compositing services is research focus. Researchers always put emphasis on service description and decidability of logical reasoning. After all, it is truly a high complicated logical to describe logically the process event in reality. As a result, people began to consider the models and methods in combination with other semantic model to represent the composition of services. FSA [6] is one of them. Calculating both structural and semantic similarity of FSA model can be a way of matching user's request and compositing services. Besides, Petri net [7] and BPEL/BPEL4WS [8] is also a nice choice to describe services process.

Semantic model of service

Considering that adding processes information to the service description can improve the accuracy of the service matching, this paper proposed Advanced Profile Model (APM), a novel service semantic description model. It can be expressed as $APM = \{BI, I, O, P, E, PI\}$, which BI is the basic information of service, including the service's name, URI and provider. I is input of service and O is output, they all record the information of data type, constrain and mapping relation between input/output parameters and ontology concepts. P is the precondition of service and E is the effect of service, they are in the form of logic expression. PI is the process information of service. If the service is atom service, PI is the concept or keywords set corresponds to transformation. If the service is Composited service, PI is the graph information extracted from FSA structure.

Atom Service Process Represent Model(ASPRM) . ASPRM can be expressed as $PI = \{S_I, S_t, S_o\}$, in which the S_I is the concept or instance or keyword set which input parameters mapped to. S_o is the concept or instance or keyword set which output parameters mapped to. the process model of service whose function is calculating the distance between two points on map can be expressed as fig 1.

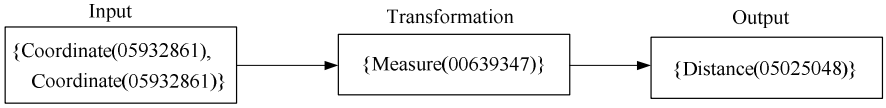

Figure 1 An example of ASPRM.

In fig 1, the input is the coordination of the two points, the number in brackets is the ID of concept, the "measure" is the description of Transformation, and the "distance" is the concept that the output parameter mapped to. Take the process description into account based on IOPE can recognize different similar service more effectively.

Composite Service Process Represent Model(CSPRM). We proposed an improve FSA model, CSPRM, based on [6], which is an enhanced service semantic description model. CSPRM can be expressed as formula: $\{Q,\Sigma,\delta,q_0,F\}$, in the formula, Q is the state set, in the set the state is set of concepts. Σ is the action alphabet that can change the state. δ is the transformation table, when given an input and an output, the state change by looking up the δ. q_0 is the start state and F is the end state. A composited service can be transferred to a process started with q_0 and ended with F. Here is an example, when given the position data, Cadmium content data of heavy metal Cadmium sampling point and PH value of Beijing farmland soil, the service is required to evaluate the level of Cadmium pollution according to "GB15618-1995 standard of soil environment quality" and display the result on map. This process can represent by the figure 2.

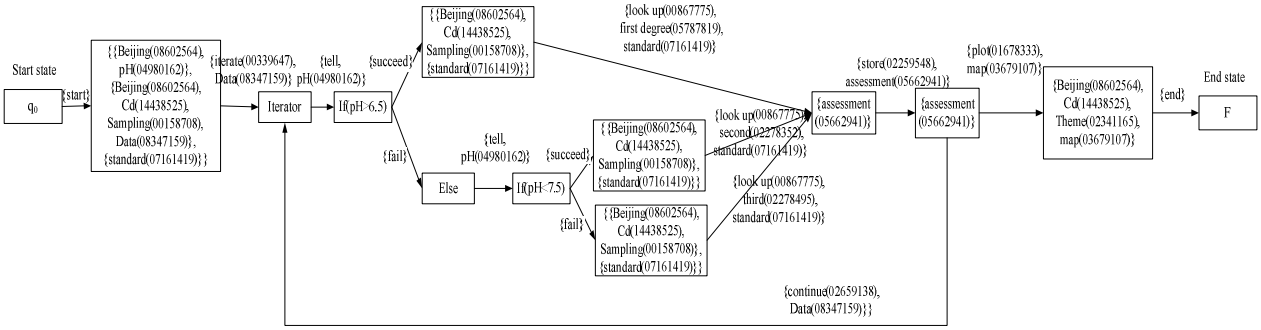

Figure 2 An example of CSPRM.

CSPRM model represents the service process by transferring the FSA form to a graphical structure model, and provides basis for the services process similarity computing.

Service Matching Algorithm

The idea of the proposed service matching method is to consider the service IOPE similarity on the basis of calculating the similarity of service processes to improve the matching accuracy of service. Typically, the service's IOPE match with the background ontology to compute similarity, this method also uses generic dictionary ontology WordNet, the proposed service matching similarity calculation method, the main consideration of the similarity computation of the four areas: similarity between concepts, similarity between the words, verb concept similarity and process similarity. The following describes the specific similarity calculation method.

Semantic Similarity of Noun Concept. In this paper, two concepts c1, c2 is set by multi-factor calculation method, which is consider the two concepts in WordNet, the shortest connection path length, information content of the most recent common ancestor. Calculated as follows:

$$Sim(c_1,c_2)=\max\{Psim(c_1,c_2),ICsim(c_1,c_2)\} \tag{1}$$

$$Psim(c_1,c_2)=\frac{depth_o-dis(c_1,c_2)}{depth_o} \tag{2}$$

$$ICsim(c_1,c_2)=\frac{IC(c_1)+IC(c_2)-2IC(NCA)}{IC(c_1)+IC(c_2)} \tag{3}$$

Formula (1) $Psim(c_1,c_2)$ is shortest path between two concepts of similarity, $ICsim(c_1,c_2)$ indicating similarity between the concepts expressed by using information content. In Formula (2), $depth_o$ represents the depth of ontology conceptual level, under the premise that ontology will not be extended, it is a constant. $dis(c_1,c_2)$ indicates the length of the shortest path between two concepts. In Equation (3) $IC(c_1)$ and $IC(c_2)$, respectively, is information content of c_1 and c_2, and NCA is the most recent common ancestor in WordNet, $IC(NCA)$ is the information content of NCA.

Semantic Similarity of Keywords. Given two words, their corresponding concept which has highest similarity between the concepts is the similarity between the two key words. Calculated by formula (4), the similarity between them:

$$Sim(w_1, w_2) = \max_{w_1 \in c_1, w_2 \in c_2} \{sim(c_1, c_2)\} \tag{4}$$

A. Semantic Similarity of Verb Concept

Because the organizational structure of WordNet, verb concepts and noun concepts are different, so the verb concept cannot form a class tree structure similar to the concepts, only part of the verb concept has directly connected with others. As a result, measure the similarity of the verb concept can not use the information content of verb concept itself. Calculating verb similarity mainly considers three factors: the length of the connection path between the concepts, the overlap of concept's gloss, the existence of the intersection between concepts which has been stemming. Given two verb concepts c_{v1} and c_{v2}, the similarity calculation formula as follows:

$$Sim(c_{v1}, c_{v2}) = \max\{Psim(c_{v1}, c_{v2}), Glsim(c_{v1}, c_{v2}), RNsim(c_{v1}, c_{v2})\} \tag{5}$$

Among them, $Psim(c_{v1}, c_{v2})$ is the connection path similarity between c_{v1} and c_{v2}. If there is no path connected, then the similarity is 0. $Glsim(c_{v1}, c_{v2})$ represents the similarity of the gloss of concept. $RNsim(c_{v1}, c_{v2})$ denotes the similarity between the verb concept of term expansion by Related-To Relationship. The formula of three kinds of similarity were calculated for:

$$Psim(c_{v1}, c_{v2}) = \frac{depth_p - dis(c_{v1}, c_{v2})}{depth_p} \tag{6}$$

$$Glsim(c_{v1}, c_{v2}) = \frac{\sum_{c_{vi} \in G(c_{v1}) \cap G(c_{v2}), c_{vi} \notin F} IC(c_{vi})}{\sum_{c_{vi} \in G(c_{v1}) \cup G(c_{v2}), c_{vi} \notin F} IC(c_{vi})} \tag{7}$$

$$Glsim(c_{v1}, c_{v2}) = \frac{\sum_{c_{vi} \in G(c_{v1}) \cap G(c_{v2}), c_{vi} \notin F} IC(c_{vi})}{\sum_{c_{vi} \in G(c_{v1}) \cup G(c_{v2}), c_{vi} \notin F} IC(c_{vi})} \tag{8}$$

In Equation (6), $depth_p$ indicates the total length of the path contains c_{v1} and c_{v2}. $dis(c_{v1}, c_{v2})$ indicates the connection path length between the two verbs. In Equation (7), $G(c_{vi})$ denotes the collection of verb in gloss of c_{vi}. F is a collection of predetermined concepts whose frequency is too high. In Formula (8), $RN(c_{vi})$ express concepts which is extended by Related-To relationship from c_{vi}.

Process Similarity. The composited service process model can represents FSA as model graph structure and the similarity between the combination of service and user requests can be calculate by graph Levenshtein distance [9]. Given two directed graph g_1 and g_2, the Levenshtein distance between them can be measured as the least number of operation(add, replace or delete terminals and edges in graph) which transform g_1 to g_2.

$$STsim(g_1, g_2) = 1 - \frac{ED(g_1, g_2)}{\max\{L(g_1), L(g_2)\}} \tag{9}$$

$$ED(g_1, g_2) = ED(V_1, V_2) + ED(E_1, E_2) \tag{10}$$

$ED(g_1, g_2)$ is Levenshtein distance of two graphical structure g_1 and g_2. $L(g_i)$ is the number of terminals and edges no repeated in the i-th graph structure. $ED(V_1, V_2)$ expresses Levenshtein distance of endpoint collection between two graphs, $ED(E_1, E_2)$ indicates that the Levenshtein distance of set of edges between two graphs.

Experiment

In this paper, several test data sets were used to observe and verify the performance of Multi factor services matching method. This article selects OWLS-TC3, test data set provide by Annual International Contest. It gives related marks of 29 in gradual and binary method. This paper uses

text similarity, structural precision and multi-factor proposed above to match 29 queries and returns the top 100 record, calculating similarity and recall rate. Text similarity method use TF / IDF keyword index to establish the service documentation, and then calculate the cosine as the similarity of the keyword vector. The structural similarity methods are calculated using the formula (9). The figure below shows the precision and recall rate diagram (P-R diagram) of the combined effect according to value of the OWLS-TC3.

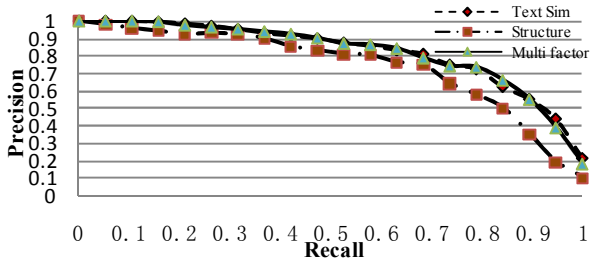

Figure 3 P-R graph of three matching method on OWLS-TC3.

It can be seen from the figure that on accuracy and recall rate, Multi factor method is partially better than Text-similarity methods. It is because that Text-similarity method using keyword-based document analysis technology, its accuracy depends on the quality of the keyword that created queries and Services Description while performance of Multi factor and the Structure depends largely on the quality of the reference ontology. In the experiment, the reference ontology of these two methods is provided by OWLS-TC3 data set. These ontology are established based on the IS-A relationship which is also relatively simple and affect the performance of the method.

Acknowledgment

This work was supported by the Postdoctoral Science Foundation of Beijing Academy of Agriculture and Forestry Sciences.

References

[1] Stroulia, E. and Y. Wang, Structural and semantic matching for assessing web-service similarity. International Journal of Cooperative Information Systems, 2005. 14: p. 407-437.

[2] Paolucci, M., et al., Semantic Matching of Web Services Capabilities, in The Semantic Web — ISWC 2002. 2002, Springer Berlin / Heidelberg. p. 333-347.

[3] Luo, A., et al. Multi-level Semantic matching of Geospatial Web Services. in The International Conference On Geoinformatics. 2009. p. 1 -5.

[4] Namgoong, H., et al. Effective semantic Web services discovery using usability. in Advanced Communication Technology, ICACT The Th International. 2006. p. 2199-2203.

[5] Jiang, Z., et al. Dynamic Description Logic Based Services Semantic Matching. 2008. p. 229 - 234.

[6] G U Nay, A. and P. Yolum. Structural and semantic similarity metrics for web service matchmaking. in EC-Web'07. 2007. Berlin, Heidelberg: Springer-Verlag. p. 129-138.

[7] Ehrig, M., A. Koschmider and A. Oberweis. Measuring similarity between semantic business process models. in APCCM '07. 2007. Darlinghurst, Australia: Australian Computer Society, Inc. p. 71--80.

[8] Sycara, K., et al., Automated discovery, interaction and composition of Semantic Web services. Web Semantics: Science, Services and Agents on the World Wide Web, 2003. 1(1): p. 27 - 46.

[9] Bunke, H., On a relation between graph edit distance and maximum common subgraph. Pattern Recognition Letters, 1997. 18(8): p. 689-694.