Using Similarity Measures for an Efficient Business Information-Exchange

Andreia Malucelli
FEUP-LIACC/ PUCPR
PUCPR – R. Imaculada Conceição, 1155
80215-901 Curitiba PR, Brazil
malu@ppgia.pucpr.br

Eugénio Oliveira
FEUP-LIACC
FEUP – R. Dr. Roberto Frias
4200-465 Porto, Portugal
eco@fe.up.pt

Abstract

Several problems are involved in the Virtual Enterprise (VE) formation process. One of the most important problems is the lack of understanding that may arise during agents' interaction, due to both the structural and semantic concepts representation heterogeneity. In the VE life cycle identification of needs, for example, it is necessary to describe the needed product or service in a way that it can be understood by all the participants. The easier way of solving this problem is to use either a common ontology or a shared one which may be understood by all the enterprise delegate agents participating in the process. However, each agent may have one of the existing different ontologies and a shared ontology will not be universal. Thus, the enterprises will not waste time converting all the content of their ontology if the new one is not considered a universal one. Due to these facts we have created an Ontology-based Service Agent which finds correspondence (similarity) between the concepts (products) of two ontologies through their respective concept's names, characteristics, relations and concept's descriptions.

1. Introduction

In a decentralized and distributed approach, interoperability refers to the way we communicate with people and software agents, the problem which hampers the communication and collaboration between agents.

Our objective is to help in the interoperability problem, enhancing agents with abilities to provide services to and accept services from other agents, as well as to use these services so exchanged to enable agents to effectively negotiate together.

By making the enterprise agents interoperable, we enable them to meet the basic requirement for multilateral cooperation. In real-life situations, real problems involve heterogeneity. This kind of problems makes the negotiation process difficult for the VE partners' selection and for the cooperation process in the VE.

The Foundation for Physical Intelligent Agents (FIPA) has analyzed the interoperability problems in heterogeneous multi-agent system (MAS) and has proposed an Ontology Agent (OA) for MAS platforms.

We have created an Ontology-based Services Agent (OSAg) [2], which is responsible for providing services to other agents in order to ensure an effective, meaningful negotiation. The OSAg provides the following services: (i) matching terms service, (ii) currency conversion service, (iii) measurement conversion service.

The matching terms service is required when some of the agents does not understand the content of a message; i.e. the item under negotiation. This service is the most complex one and it is implemented based on lexical and semantic similarity measures. The lexical measures are used to compare attributes, relations between concepts and descriptions of the concepts.

We have classified attributes according to their data value types and considered the relation has-part. Moreover, the Leacock-Chodorow (LCH) method [1] based on WordNet [8] is applied between concept names.

The currency conversion service may be useful in the calculation of prices when agents are dealing with different currencies. Furthermore, currency conversion service is implemented as a Web Service.

Similarly, the measurement conversion service may be useful when agents are dealing with different measure units. In addition, the ontology editor Protégé [10] is integrated in the framework to facilitate the creation and maintenance of ontologies.

This paper focus on the matching terms service. The next section presents the architecture of the proposed system. In Section 3 the matching terms service is detailed. Section 4 points out some experiments and Section 5 outlines some related works. Finally, we present some conclusions in Section 6.

2. System Architecture

This framework includes 4 types of agents: facilitator agent, enterprise agents (good/product/services suppliers and customer), and ontology-based services agent. Each enterprise agent has its own architecture and functionalities (some developer will design and build the ontology with some tool and, later, the agent will access the generated file/database).

The Facilitator Agent (FAg) is the entity that matches the right agents and supports the negotiation process.

Customer Enterprise Agents (CEAg) represent enterprises interested in buying components to build a final product. Several suppliers in the world may have these components with different prices and conditions. Each CEAg sends a message to the facilitator announcing which composed product/service is needed.

Supplier Enterprise Agents (SEAg) represent enterprises interested in providing some kind of good/product/service. Whenever there is a needed product, the facilitator agent conveys this announcement to all registered interested supplier enterprise agents.

Ontology-based Services Agent (OSAg) keeps monitoring the whole conversation providing services, trying to help when some message is not fully understood by some of the participants.

3. Matching Terms Service

Similarity evaluations among ontologies may be achieved if their concepts' representations share some components. If two ontologies use at least one common component (attributes, relations, hierarchy, types) then they may be compared. Usually characteristics (attributes) provide the opportunity to capture details about concepts.

We have integrated three different similarities matching, which are performed by the OSAg in our MAS platform. Afterwards, a final result is calculated in order to make a statement if the compared concepts (products, in the context of this work) have the same meaning.

The three methods are: (i) calculating an n-grams [11] value for the attributes and relations of the concepts; (ii) calculating an n-grams value for the description of the concepts, and (iii) applying the LCH method based on WordNet to detect semantic similarity between both concepts.

3.1. Similarity Matching between Attributes

This methodology considers the attributes (characteristics) attached to a concept. It is based on the assumption that if two concepts in the same domain describe the same product, there is high probability of having a set of mandatory characteristics for the product specification. So, these similar characteristics, or at least similar written, occur when describing the same concepts.

However, before comparing the attributes, they are classified according to their data types. String, float, integer, boolean and the relation has-part are considered separately for the comparisons, in order to reach a more reliable result. This classification reflects the fact that attributes, which describe the same concept, are in most cases of the same type.

Applying some similarity measure method without splitting up the attributes could lead to a result that could be misinterpreted. This method is detailed in [2].

3.2. Similarity Matching between Descriptions

Usually, in a specific domain, when experts describe concepts defined in an ontology, they use common, technical words, so that it might be possible to detect these similarities in the descriptions.

A previous process is necessary in order to remove the stopwords, since the most representative words have to be extracted from the description. This process leads to a short, precise description of the product. Words that occur several times in one description are only considered once, and punctuation marks are eliminated as well. After this previous process, an n-grams matrix is calculated between each word in the descriptions. The highest value for each line is taken and the $r_{n\text{-grams}}$, [2] is calculated, where *n* is the number of words in the description of the requested product.

3.3. Similarity Matching between Concepts

We are using a CPAN module [5] that implements a variety of semantic similarity measures that can be used in conjunction with WordNet. In particular, we have evaluated the measures [1] of Resnik, Jiang-Conrath, Leacock-Chodorow, Hirst-St.Onge, and Wu-Palmer in order to select the one which fits the best to our case. The Leacock-Chodorow (LCH) showed to be the best for our case. The LCH method requires two input parameters in the format "word#pos#sense".

Several taxonomies exist inside WordNet, but in the underlying scenario only nouns are relevant. Since a word in WordNet has several senses, the relevant sense has to be chosen.

The proposed solution takes all senses into account. Each sense of one concept is compared with each sense of the other one. The highest value of all these combinations is considered as similarity value. This does not guarantee that the right sense is always chosen. However, experiments showed that this procedure gives applicable results in general since the highest sense indicates, in most cases, that two concepts in the same domain have been compared.

A disadvantage of this approach is the high number of comparisons: if both words have several senses it leads to a huge amount of comparisons. The results are stored on the server side and will be available in the next negotiation round. This way the whole process does not have to be repeated if the same concepts have to be compared again.

3.4. Final Matching

If at least two of the three matching explained before deliver a result, the system makes a statement concerning the correspondence of terms.

3.4.1 Weighting. The idea of using a weighting for computing a final result is based on the assumption that the single results delivered by the comparisons do not have the same reliability. As LCH is based on WordNet it delivers the most trustworthy result. The other comparisons depend more on the arbitrariness of ontology developers.

In order to reflect these assumptions, different weightings for the single results are added and considered for the calculation of the final result, and the formula (1) is used. The *n* can be 2 or 3, depending on the number of single results. The values for weighting were established after several experiments.

$$sim_{term1/term2} = \sum_{i=1}^{n} result_{methods} * weighting \Big/ n \qquad \textbf{(1)}$$

Using weighting, the comparison results are higher in general, not only the ones for terms that describe the same product, but also the values for terms that do not have any similarity.

3.4.2 Classification. The calculated final results range from 0 to 1, where 0 indicates no similarity at all and 1 indicates 100% matching. In order to define the threshold, the measures precision and recall were used. Until a threshold of 0.7 the applied methods have retrieved more than 70% of all information, and the information is still 100% precise. Whether the threshold is decreasing, precision is also decreasing. The point in which the precision and the recall overlap is approximately at a threshold value of 0.55. This determines the threshold used by the application. Lower values do not indicate enough similarity. If more than one comparison result is above the threshold 0.55, the item with the highest value is proposed. Since the results of proposed values range from 0.55 to 1, a classification for these values is used. The different levels of correspondence are: weakly-matching (threshold between 0.55 and 0.59), approximately matching (threshold between 0.6 and 0.69) or strongly-matching (threshold between 0.7 and 1.0). The classification is attached to the message sent by the OSAg to the SEAg when proposing a matching concept.

4. Experiments

A scenario demonstrating the utility of as well as the problems with VE, negotiations, and ontology-based services has been selected. The scenario uses the cars' assembling domain. This is a suitable scenario because it involves several services suppliers' enterprises and consequently several different negotiations.

We have built two different ontologies using the ontology-building tool Protégé, although in the same domain. Each one of the agents (SEAg and CEAg) used one of these different built ontologies. The ontology used by the CEAg had 27 concepts and the one used by the SEAg had 26 concepts. The ontologies specifications include a concept (item/product), its characteristics (attributes) with the correspondent data types, a natural language description explaining the meaning of the concept, and a set of relationships among these concepts.

For the experiments, the CEAg sequentially asked for all the items listed in its ontology, when the SEAg was not able to understand the requested item, it asked the OSAg for help. The OSAg performed its tasks accordingly and returned both the final result and the classification to the SEAg. Consequently, the agents were able to a more efficient business information exchange.

The SEAg's ontology had 19 concepts with correspondents in the CEAg's ontology, represented in a different way (concept names, attributes, relations and descriptions). There were also 8 concepts without correspondence, probably the SEAg did not provide those items. When calculating the similarity between all the concepts using the established threshold 0.55, the OSAg could find 17 correct concepts out of the 19 right matching ones. The OSAg did not return 2 concepts that had a matching and the other 8 concepts could not be returned because there was no matching.

The proposed method achieved 89% of accuracy. Nevertheless if we consider the not-established matching due to the lack of a correspondent concept, the accuracy rises to 92%. The efficiency of the method depends on the quantity of available information and the quality of the concepts description. Obviously, the item has to be in the WordNet database.

5. Related Work

Different approaches have investigated the interoperability problem in MAS. However, they did not indicate how agents would dynamically interact. Most of the proposed approaches to solve the interoperability problem in MAS take into account that both ontologies are known or shared. In our context, each enterprise agent keeps its private ontology and just sends the needed information about it to enable the OSAg help.

An implementation of [7] is presented in [9]. It is a sample application of an ontology shopping service that integrates multiple database schemata to verify and demonstrate the specification. However, there is no possible way to match terms between ontologies.

[3] proposes the use of B2B agents to manage the automated composition of required Web services stored

in registries. In this approach agents create small local consensus ontology to facilitate the discovery and understanding. It involves syntactic and semantic equivalence as well as the use of wordnet, taking into consideration only the concepts.

[4] presents an upper ontology based on content reference model that provides the semantic for message content expressions. A tool is also presented which can assist agent programmers in designing message content ontologies with the Protégé.

[6] proposes a domain independent method for handling interoperability problems by learning a mapping between ontologies. The learning method is based on exchanging instances of concepts defined in the ontologies.

[12] describes an approach to ontology negotiation that allows web-based information agents to resolve mismatches in real time without human intervention. The negotiation process culminates in one or both agents modifying their ontology to introduce a new concept, a new distinction, or simply a new term for an existing concept.

6. Conclusions and Future Work

In this paper we have addressed a solution for the interoperability problem in MAS and presented our solution. The proposal includes the use of lexical and semantic similarity measures to facilitate the interoperability.

The ontology-based service agent finds correspondence (similarity) between the concepts (products) of two ontologies through their respective concept names, characteristics, relations and concept descriptions, without needing a complex process of ontologies integration. Agents may have their ontology built using different ontology tools and stored in different ways.

We aim to test our approach with real ontologies in the car's assembly domain, even so we have built some ontologies based on real information, however it would be interesting to observe the methods performance in a real scenario. Moreover, the ontology-based service agent is being enhanced with learning characteristics, thus the OSAg would be able to learn the concepts already compared, this way avoiding to perform all the similarity matching process in the next negotiation round when the same item is requested.

7. References

[1] A. Budanitsky, G. Hirst, "Semantic distance in WordNet: An experimental, application-oriented evaluation of five measures", In NAACL Workshop on WordNet and Other Lexical Resources, 2001, Pittsburgh, USA, 2001.

[2] A. Malucelli, D. Palzer, and E. Oliveira, "Combining Ontologies and Agents to help in Solving the Heterogeneity Problem in E-Commerce Negotiations". In DEEC. IEEE., Tokyo, Japan, 2005.

[3] A. Williams, A. Padmanabhan, and M.B. Blake, "Local Consensus Ontologies for B2B-Oriented Service Composition". AAMAS, Eds. J. Rosenschein, T. Sandholm, M. Wooldridge, M. Yokoo, ACM Press, Melborne, Australia, 2003, pp. 647-654.

[4] C. van Aart, R. Pels, G. Caire, and F. Bergenti, "Creating and Using Ontologies in Agent Communication", Workshop on Ontologies in Agent Systems, AAMAS, Bologna, Italy, 2002.

[5] CPAN-Module, WordNet::Similarity, at http://www.d.umn.edu/~tpederse/similarity.html, 2004

[6] F. Wiesman, N. Roos, "Domain independent learning of ontology mappings", AAMAS, Eds. Nicholas Jennings, Carles Sierra, Liz Sonenbergm, Milind Tamble, ACM Press, New York, USA, 2004, pp. 846-853.

[7] FIPA-OSS, Ontology Service Specification, at http://www.fipa.org/specs/fipa00086, 2004

[8] G. Miller, "WordNet:A Lexical Database for English", Communication of ACM, 38(11), 1995, pp. 39-41.

[9] H. Suguri, E. Kodama, M. Miyazaki, H. Nunokawa, and S. Noguchi, "Implementation of FIPA Ontology Service", In Proceedings of the Workshop on Ontologies in Agent Systems, AAMAS, Montreal, Canada, 2001.

[10] J. Gennari, M.A. Musen, R.W. Fergerson, W.E. Grosso, M. Crubézy, H. Eriksson, N.F. Noy, and S.W. Tu, "The Evolution of Protégé: An Environment for Knowledge-Based Systems Development", Technical Report, SMI Report Number: SMI-2002-0943, 2002.

[11] M. Damashek, "Gauging Similarity via N-Grams: Language-independent Sorting, Categorization, and Retrieval of Text", Science 267, 1995, pp. 843-848.

[12] S.C. Bailin, W. Truszkowski, "Ontology Negotiation between Scientific Archives". In SSDBM'01, Fairfax, Virginia, 2001.