A Semantic Similarity Model for Mapping Between Evolving Geospatial Data Cubes

Mohamed Bakillah, Mir Abolfazl Mostafavi, and Yvan Bédard

Chaire Industrielle CRSNG en bases de données géospatiales décisionnelles
Centre de Recherche en Géomatique, 0611 Pavillon Casault, Département des Sciences Géomatiques Université Laval, Québec, Canada, G1K 7P4
Mohamed.bakillah.1@ulaval.ca
Mir-Abolfazl.Mostafavi@scg.ulaval.ca
Yvan.Bédard@scg.ulaval.ca

Abstract. In a decision-making context, multidimensional geospatial databases are very important. They often represent data coming from heterogeneous and evolving sources. Evolution of multidimensional structures makes difficult, even impossible answering to temporal queries, because of the lack of relationships between different versions of spatial cubes created at different time. This paper proposes a semantic similarity model redefined from a model applied in the ontological field to establish semantic relations between data cubes. The proposed model integrates several types of similarity components adapted to different hierarchical levels of dimensions in multidimensional databases and also integrates similarity between features of concepts. The proposed model has been applied to a set of specifications from different inventory in Montmorency Forest in Canada. Results show that the proposed model improves precision and recall compared to the original model. Finally, further investigation is suggested in order to integrate the proposed model to SOLAP tools as future works.

Keywords: Semantic similarity models, ontology, mapping between ontologies, geospatial data cubes.

1 Introduction

OLAP tools and their spatial extension SOLAP, which are based on multidimensional structures, were introduced to support analysis in a decision-making context and to allow users to easily access and explore the data according to various perspectives [1]. Multidimensional structures are composed of dimensions, measures and facts. Dimensions are the analysis themes and can be spatial, typically having levels described by geometric objects, such as polygons, for cartographic representations. Measures are the numerical attributes analyzed against different dimensions. Facts express the value of measures with respect to a specific combination of dimensions members for different aggregation levels. For example, the multidimensional structure can be made of the spatial dimension geography, formed of levels *city* < *area* < *country*, and measure *birth rate*. Multidimensional structure is brought to undergo evolution, affecting facts but also structure of dimensions [2], which are usually considered as static. Dimension

R. Meersman, Z. Tari, P. Herrero et al. (Eds.): OTM Workshops 2006, LNCS 4278, pp. 1658 – 1669, 2006.

members can also be affected by semantic evolutions. An example of such is a modification resulting from the changes to a regulation affecting the management of a given territory. Such evolution of multidimensional structures affects temporal queries, and may lead to false results, when result there is [3] [2] [4]. Existing approaches suggested to manage the evolution in multidimensional data cubes [5] [2] [6] [7] do not explicitly consider semantic evolution as they manage only the case of explicit data evolution, i.e. when evolution is realized by evolution operators (add, delete, merge members of schema of instances of dimension, etc.) Evolution can also happen when several multidimensional databases represent the same reality for different epochs, for example when the data are collected independently for the same territory every 10 years as in forestry in Quebec. In this case, relations between the seemingly similar databases are difficult to establish and the evolution problem is more complex to solve since we need to restore these relations to answer temporal queries. In order to answer this latter problem, this paper proposes a new approach of semantic mapping between data cubes. The proposed approach is based on an ontological approach for the assessment of semantic relations between members of different data cubes. Multidimensional structures of a same area and their metadata can be seen as an evolving ontology and thus a semantic similarity model can be used as a mapping function. This similarity model is adapted to complex data and is flexible enough to support several data types. It defines a specific measure of similarity for the aggregated levels of a dimension hierarchy. Follow up on this work will show how the developed mapping function can support temporal queries processing in data cubes.

The reminder of this paper is as follows: section 2 presents a state of the art on ontology mapping and semantic similarity models. Section 3 describes the proposed approach and the similarity model used as a mapping function between data cubes. Section 4 shows the application and evaluation of the proposed approach in a forestry context. In section 5, we conclude this article.

2 State of Art on Ontology and Semantic Similarity Models

A suitable approach to overcome problems of semantic heterogeneity and evolution lies in ontologies, which are specifically designed to represent semantics and knowledge about data. In AI (Artificial Intelligence), ontology is defined as an explicit specification of a conceptualization [8]. In other words, an ontology is the outcome of a conceptual modeling process. Generally, the taxonomic structure of an ontology forms a graph where nodes represent concepts and arcs represent relations between them. An ontology thus constitutes an interesting framework for the discovery of semantic relationships between concepts, upon which we have founded our approach, i.e. a multidimensional structure with metadata can be regarded as an ontology. Moreover, just like the multidimensional structure, ontologies evolve, following modifications of specifications, standards, definitions of concepts, etc. [9] [10] [11].Ontology mapping aims at establishing relationships between ontologies while preserving their own structure. Among the approaches of ontology mapping, some use a semantic similarity model to relate concepts [12] [13] [14]. Similarity models can be classified according to the representation of the concepts they use: graphs, features of concepts, information content-based models, vector space models or a combination of different types of models (hybrid models).

Similarity models using ontology graphs are based on the assumption that a hierarchy of concepts is organized according to semantic similarity lines. Consequently, concepts are similar if the distance which separates them in the graph is short, the distance being given by the shortest path along the arcs to join both concepts [15]. The distance between two nodes is not necessarily uniform, thus other models take into account local density of nodes, depth of concepts in the graph, total depth of the graph and the force of the relations [16] [17] [18]. Models using features of concepts, based on set theory, are founded on the comparison of sets of features describing concepts. The ratio model [19], also used in other approaches [20] [21], compares the intersection set (common features) to the sets of exclusive features of each concept.

Models based on information content stipulate that the similarity between two concepts is getting higher as the shared information content increases. Information content of a concept is a logarithmic function of the probability of its components to appear in the ontology. Similarity between two concepts is given by the information content they share, i.e. by the information content of the first common parent in taxonomy [22]. According to another approach, information content of a concept is a function of the number of hyponyms and of the number of concepts in taxonomy [23]. Vector space models use the analogy where semantic proximity between concepts is represented by proximity in a vector space. This model is mainly used in information systems to represent documents, although it is also used for concepts in geometrical similarity models [24] [25], where those are represented by points or regions in a multidimensional conceptual space. The semantic distance between concepts is given by a metric, such as the Minkowski distance or cosine [26] in information systems.

Finally, hybrid models, such as the Matching Distance model, merge properties of several models into one [20] [27] [25]. Matching Distance model [20] is based on the ratio model [19], integrates context and distance in ontology graph and was designed to associate spatial entity classes from different ontologies. In Matching Distance Model, global similarity is a weighted sum of similarity between the different types of features (attributes, parts and roles) of concepts, lexical similarity (name of concepts) and neighborhood similarity in the graphs of ontologies:

$$S_g(c_1,c_2) = \omega_l S_l(c_1,c_2) + \omega_c(\omega_a S_a(c_1,c_2) + \omega_p S_p(c_1,c_2) + \omega_f S_f(c_1,c_2)) + \omega_n S_n(c_1,c_2) \qquad (1)$$

where the different similarities are given by an adaptation of the ratio model (shown in next section) and neighborhood similarity by :

$$S_n(c_1,c_2) = \frac{N_1 \cap N_2}{N_1 \cap N_2 + \alpha(c_1,c_2) * \delta(c_1, N_1 \cap N_2) + (1-\alpha(c_1,c_2)) * \delta(c_2, N_1 \cap N_2)} \qquad (2)$$

N1 and N2 are entities forming the neighborhood of concepts in the ontology graph, α is the distance between concepts in graph and δ is the difference function between neighborhoods of concepts. This model has the advantage of being complete and to take into account the maximum of information contained in ontology, compared to other models. However, it only considers features as words. This can be insufficient since spatial entity may be defined by more complex features such as domain values or texts. Also, it does not consider the degree of similarity between features of concepts, nor the specificity of concepts from aggregated levels of graph. Based on this model, in this article we propose a new semantic similarity model that overcomes these limitations and that can be used as a mapping function between members of the schema of instances of dimension.

3 Proposed Approach

The multidimensional structure and its metadata are considered as the ontology of the databases in order to evaluate semantic similarity between members of the schema of instances of dimensions in different cubes. First, the user defines the context which represents a set of concepts that share a feature of interest (for example *ecological zone*) and will be used to compute weights for different similarities. Similarity is then evaluated on three levels: between features of concepts, between concepts of the detailed level (finest level of granularity in the schema of instances of dimension) and between concepts of aggregated levels. The global similarity allows computing the matrices of mapping which relate concepts from two levels of hierarchy of different cube versions. The proposed approach is shown on figure 1.

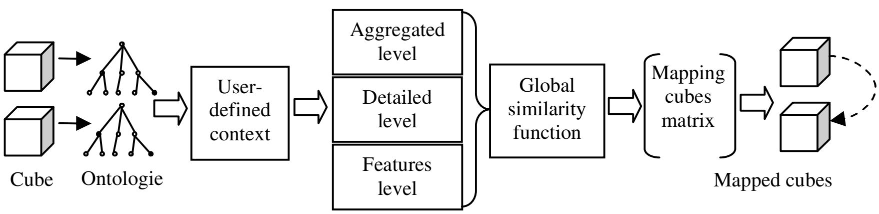

Fig. 1. Global approach for semantic mapping between evolving data cubes

Weights for the different similarities are defined by computing relevance of each type of features using the principles of commonality or variability. Relevance following the commonality principle (P_t^{comm}) is defined as the sum of the number of occurrences o_i of each feature i of type t in the concepts definitions, divided by the number n of features in the context, while relevance computed with using the variability principle (P_t^{var}) is defined as the converse of commonality principle. Weights are defined as following:

$$\omega_t = \frac{P_t}{\sum_i P_i} \quad \text{i=type of feature where } P_t^{comm} = \frac{1}{n}\sum_{i=1}^{n} o_i \quad \text{or} \quad P_t^{var} = 1 - \frac{1}{n}\sum_{i=1}^{n} o_i \qquad (3)$$

3.1 Theoretical Framework

This section introduces basic definitions of our approach. First, we define the ontology and concepts, and then the mapping function and matrix of semantic mapping that relate concepts of different versions of the ontology. The mapping function is based on the similarity model that will be defined in the next section.

Definition of Concepts. A concept c of a version O^i is defined such as c = {id_c, name_c, P, D, L_D, O^i, t}, where id_c is the concept identifier, name_c is its name, P is a set of features (attributes, parts and functions) of the concept, D and L_D are respectively the dimension and the hierarchical level to which the concept belongs and t is a valid time interval for concept c. The set of features P may contain features which domain values are text or intervals given by specifications on concepts.

Mapping Function. This ontology-like representation of the multidimensional structure forms a framework for discovering semantic relations between members of the different cubes. Semantic relations are established by a mapping function which quantifies, by means of a semantic similarity measure, a similarity relation between two concepts. This mapping function takes two forms, depending on whether concepts belong to the detailed level or an aggregated level of a hierarchy. The mapping function for the detailed level relates concepts c_1^i and c_2^j of the detailed levels $L_d^{\ i}$ and $L_d^{\ j}$ of two versions i and j of ontology with the similarity measure S_{gd} :

$$f_d : S_{gd}(c_1^i, c_2^j) \text{ with } c_1^i \in L_d^i \text{ et } c_2^j \in L_d^j \tag{4}$$

The mapping function of the first aggregated level relates either a concept of the detailed level with a concept of an aggregated level or two concepts from aggregated levels with the similarity measure S_{g_agg}. The S_{g_agg} similarity measure depends on the similarity S_{gd} between the components of the concepts:

$$f_{agg1} : S_{g\_agg}(S_{gd}(P_1^i, P_2^j)) \quad \text{with } P_1^i = \begin{cases} \text{components of } c_1^i & \text{if } c_1^i \in L_{agg}^i \\ c_1^i & \text{if } c_1^i \in L_d^i \end{cases} \tag{5}$$

Components of a concept are children in the hierarchy. We can generalize the mapping function for arbitrary aggregated level *n* using the recursive principle:

$$f_{agg\_n} : S_{g\_agg\_n}(S_{g\_agg\_n-1}(S_{g\_agg\_n-2}(...(S_{gd}(P_1^i, P_2^j))))) \tag{6}$$

Matrix of Semantic Mapping. The mapping function allows computing the elements of the matrix of semantic mapping which relate concepts from two levels of two versions O^i and O^j of the ontology. We define a matrix of mapping for each combination of level of the two versions. Let H (D, O^i) = {$c_1^{\ i}$, $c_2^{\ i}$ …, $c_k^{\ i}$ … $c_n^{\ i}$} be the set of n concepts forming the level L_1 of dimension D of the version O^i and H (D, O^j) = {$c_1^{\ j}$, $c_2^{\ j}$ …, $c_l^{\ j}$ … $c_m^{\ j}$} be the set of m concepts forming level L_2 of dimension D of the version O^j. The matrix of semantic mapping M as dimension card(H (D, O^i))×card(H (D, O^j)), where card() represents the number of element in a set. Each of its elements is defined by $M(D,O^i,O^j)_{kl} = f(c_k^{\ i}, c_l^{\ j})$ where f is the mapping function for the detailed level f_d if L_1 and L_2 are detailed levels and f_{agg} if L_1 or L_2 is an aggregated level. Searching a semantic relation between all possible levels is necessary in order to identify concepts that may have changed level.

3.2 Redefined Semantic Similarity Model

The model described in this section improves the model Matching Distance (MD) by allowing us to measure similarity between the features of concepts and presents a new similarity measure for concepts of aggregated levels. Similarity is computed in a three step recursive process: (1) between features of compared concepts, (2) between concepts of the detailed level and (3) between concepts of aggregated levels of the schema of instance of dimensions, using the similarity values of detailed level concepts. Features can be complex and we include measures of similarity for texts and domain values (intervals).

3.2.1 Features Level Similarity

Matching Distance model (MD model) is based on the ratio model [19] which suggests that similarity is a function of the sets of common features and sets of exclusive features:

$$S(c_1, c_2) = \frac{f(C_1 \cap C_2)}{f(C_1 \cap C_2) + \alpha f(C_1 - C_2) + \beta f(C_2 - C_1)} \tag{7}$$

where is C_1 the set of features of c_1 and C_2 is the set of features of c_2, f is a monotonous increasing function and $\alpha \geq 0$, $\beta \geq 0$ are parameters which give relative importance to the sets of exclusive features. In MD model, f is the set cardinality and for a feature, being part of the set of common features is a binary function, i.e. a feature is or is not part of this set, but it cannot be included in the set in a partial way. Matching Distance model thus underestimates the similarity of two concepts, because it does not consider the degree of similarity of the features. However, we consider that, just as concepts, features may share some degree of similarity, so the contribution of a pair of features to the set of common features must be related to their percentage of similarity, so the first step of similarity assessment is to compute similarity between features. Similarity measures employed for text and intervals features are as follows:

Similarity Measure for Text. The similarity model employed to compare texts is generally used in information systems. Segmentation and indexing processes are resumed in figure 2. Result of indexing is a set of informative segments forming the lexicon.

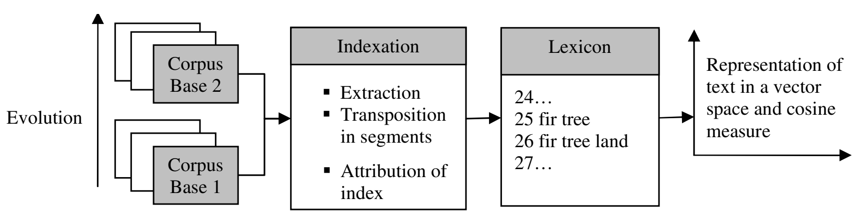

Fig. 2. Text similarity evaluation

Similarity is given by cosine measure [26], where each text is represented by a vector whose components $v_1, v_2, \ldots v_l$ are frequencies of informative segments of the lexicon :

$$sim(A_i, B_j) = \sum_{k=1}^{l} v_{ki} v_{kj} \Bigg/ \sqrt{\sum_{k=1}^{l} v_{ki}^2 \sum_{k=1}^{l} v_{kj}^2} \tag{8}$$

Similarity Measure for Intervals. Replacing summation with integration in the cosine measure (equation (8)) produces a vector space similarity measure adapted for the case of continuous data, i.e. where the number of dimensions in the vector space is infinite [28]. Frequency of terms, which usually quantifies the components of vectors representing the concepts, becomes a continuous function of density ρ(r) that indicates distribution of values in a continuous range:

$$\mathrm{sim}(c_1,c_2)=\frac{I(c_1,c_2)}{\sqrt{I(c_1,c_1)I(c_2,c_2)}} \text{ where } I(c_1,c_2)=\int \rho_1(r)\rho_2(r)dr \tag{9}$$

This similarity measure can be applied to compare intervals since they are infinite sets of values. Then, the functions of density $\rho_1(r)$ and $\rho_2(r)$ indicates the distribution of the values in the compared intervals and $I(c_1,c_2)$ represent the intersection function between both intervals. Result is a similarity value that lies between 0 and 1. These similarity measures for text and intervals are included in equation (10).

3.2.2 Detailed Level Similarity

According to the principle that the contribution of a pair of features to the set of common features must be related to their percentage of similarity and considering that $C_{1t} = \{A_1, A_2\dots A_i\dots, A_n\}$ and $C_{2t} = \{B_1, B_2\dots B_i\dots, B_m\}$ are the respective sets of feature of type t of concepts C_1 and C_2 (t=attributes, parts, or role) function of the intersection *f* defined in equation (7) can be defined in the following way:

$$f(C_{1t}\cap C_{2t})=\sum_{i=1}^{n}\sum_{j=1}^{m}\mathrm{sim}(A_i,B_j) \tag{10}$$

where $sim(A_i, B_j)$ evaluate similarity between features and depends on the data type of the feature A_i and B_j. The difference between the two sets of exclusive features is defined by:

$$f(C_{1t}-C_{2t})=card(C_{1t})-f(C_{1t}\cap C_{2t}) \text{ and } f(C_{2t}-C_{1t})=card(C_{2t})-f(C_{1t}\cap C_{2t}) \tag{11}$$

On the detailed level, global similarity is given by the following,

$$S_{gd}(c_1,c_2)=\omega_l S_l(c_1,c_2)+\omega_c(\omega_a S_a(c_1,c_2)+\omega_p S_p(c_1,c_2)+\omega_f S_f(c_1,c_2))+\omega_n S_n(c_1,c_2) \tag{12}$$

where the different similarity terms now depend on the functions defined by equations (10) to (12) which are incorporated in equation (7), where S_n is neighborhood similarity given by equation (2). The advantage of such a model is its flexibility since any type of other feature for which it is possible to define a similarity measure giving values between 0 and 1 can be incorporated in equation (10).

3.2.3 Aggregated Level Similarity

Concepts from aggregated levels are formed by the underlying concepts in the hierarchy, i.e. their subordinated concepts (the components). In some cases, MD model may be insufficient to assess similarity between concepts of aggregated level which may have no intrinsic feature. For example, when concept are spatial zones, concepts of aggregated levels are an aggregation of concepts of detailed level which are related to them by part-of relations, and thus have no features except parts. Names of concepts of spatial zones may be unmeaning for lexical similarity: for example, even if the concept *forestry station* shares no lexical commonality with the concept *landscape unit*, they represent a very close reality in forestry. Following this last remark, parts of concepts of aggregated levels may be impossible to compare directly only by their names; it is necessary to assess their features similarity in order to know to which level parts are similar. In these cases, global similarity of MD model reduces to neighborhood similarity, thus not taking into account similarity between components of concepts. To extend the MD model to these cases, the following model for similarity assessment between concepts of aggregated level is proposed. Similarity assessment for aggregated levels is a recursive process, i.e. the comparison of the sets of components of each concept is also a similarity assessment. Consider that $P_1= \{p_{11}, p_{12}\dots p_{1i}\dots, p_{1n}\}$ and $P_2 = \{p_{21}, p_{22}\dots p_{2i}\dots, p_{2m}\}$ are the sets of components of concepts c_1 and c_2 respectively. The similarity for aggregated levels between the concept c_1 and c_2 is given by:

$$S_{agg}(c_1,c_2) = \frac{f(P_1,P_2)}{f(P_1,P_2) + \alpha(c_1,c_2)D\left(P_1 - P_2\right) + \beta(c_1,c_2))D(P_2 - P_1)} \tag{13}$$

The function f represents the set of common components to concepts c_1 and c_2 and is evaluated by summing the similarities between the most similar concepts:

$$f(P_1,P_2) = \begin{cases} \sum_{k=1}^{\mathrm{card}(P_1)} \max\left[S_{gd}(p_{1k,}p_{2i})\right] & \text{if } \mathrm{card}(P_1) \leq \mathrm{card}(P_2),\ i \in \left[1,\mathrm{card}(P_2)\right] \\ \sum_{k=1}^{\mathrm{card}(P_2)} \max\left[S_{gd}(p_{1i,}p_{2k})\right] & \text{if } \mathrm{card}(P_1) \succ \mathrm{card}(P_2),\ i \in \left[1,\mathrm{card}(P_1)\right] \end{cases} \tag{14}$$

Image of function f is constrained by the cardinality of the smallest set of components, because the cardinality of intersection set must be smaller or equal to the cardinality of the smallest set: f:[0,1]×[0,1]→[0,min{card(P_1),card(P_2)}]. Differences between sets of exclusives components are given by the cardinality of components sets from which we substract the intersection function f:

$$D(P_1 - P_2) = card(P_1) - f(P_1,P_2) \quad \text{and} \quad D(P_2 - P_1) = card(P_2) - f(P_1,P_2) \tag{15}$$

Global similarity for aggregated levels is given by the following, where ω_p is the weight for similarity of aggregated level, since this last one evaluates similarity between parts (component) in hierarchy:

$$S_{g\_agg} = \omega_l S_l(c_1,c_2) + \omega_c(\omega_a S_a(c_1,c_2) + \omega_p S_{agg}(c_1,c_2) + \omega_f S_f(c_1,c_2)) + \omega_n S_n(c_1,c_2) \tag{16}$$

4 Evaluation of the Proposed Approach

The evaluation of our model has been done using a Java application with forestry spatial data and illustrates the accuracy of this redefined model as well as the increased performance of that model compared to the MD model. The data came from four different inventories (1973 to 2002) of Montmorency experimental forest of Laval University. Each inventory is associated to a data cube. A forest inventory consists in partitioning space in zones characterized by homogeneous properties of term of density, species, height, etc.(table 1), resulting in a set of basic spatial entities. Basic spatial entities are aggregated in higher level spatial entities, forming the hierarchy of the data cube spatial dimensions. For research, regulatory and environmental reasons, the specifications of theses spatial zones have changed from one inventory to another.

Table 1. Example of evolving basic spatial entities (from Montmorency forest specifications)

Year inventory	Attributes				Parts	Roles
	Age	Height	Density	...	Species	Zone type
1992	[20,40] years	[7,12] m	[61,81]%	...	Mixed zone where leafy tree represent over 50 % of ...	Ecological
2002	[30,45] years	[10,12] m	[55,81]%	...	Mixed zone where white-birches take over 45% of...	Ecological

At first, a simulation was carried out to validate the redefined semantic similarity model. Behaviour of the model was evaluated according to a sample of concepts for which the percentage of common features with a reference concept follows a linearly increasing function. Figure 3 shows that the model follows the predicted behaviour compared to the variability of the concepts.

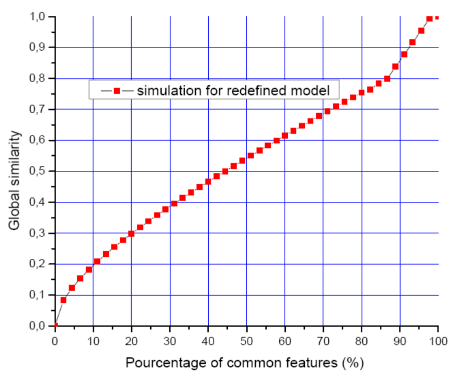

Fig. 3. Behaviour of the redefined model

Results of similarity obtained with the MD model and our redefined model were also compared, showing that for different classes of similarity, the MD model underestimates the value of the similarity because it rejects the features which are partially similar, whereas the redefined model gives values of similarity closer to the reality (table 2).

Table 2. Comparison of some values obtained for the redefined model and the MD model

Zone number	MD model	Redefined model	Expected range of values
23-3	0.3712	0.5874	Average similarity
11-3	0.1650	0.3990	Low similarity
2-40	0.1433	0.3784	Low similarity
12-54	0.0206	0.1639	Very low similarity

The efficiency of both models was compared by evaluating precision and recall, which are metrics currently used in information system:

$$\text{Precision} = \frac{\text{number of correct mapping}}{\text{number of detected mapping}} \text{ and Recall} = \frac{\text{number of correct mapping}}{\text{number of reference mapping}} \quad (17)$$

Expected mapping were manually evaluated using specifications and cartographic data for the compared spatial entities. The evaluation was carried out with 25 zones identified in the 1984 inventory and 77 zones identified in the 1992 inventory, since in 1992 the same surface was divided in smaller entities than in 1984. Results of precision-recall curves are shown on figure 4.

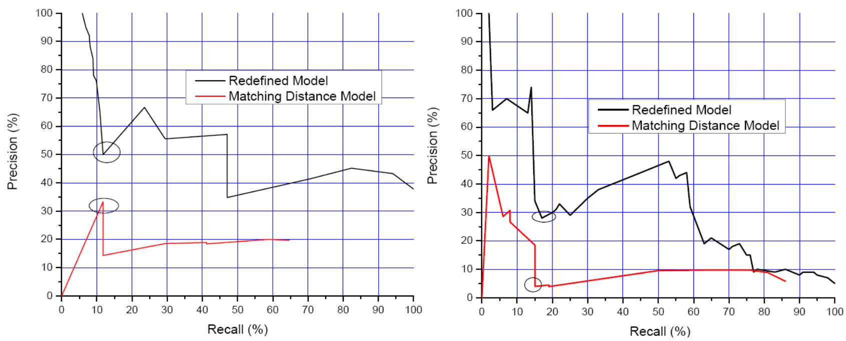

Fig. 4. Precision-recall curves for aggregated levels (left) and detailed level (right)

Values were obtained by successively applying decreasing threshold values, threshold indicating that beyond this value pairs of entities could be considered similar and part to the set of detected mapping. Set of correct mapping is a subset of the set of detect mapping and corresponds to the pairs of entities that are also part of the set of reference mapping. Results show that performance of the redefined model is higher than performance of MD model in any point, in particular, the difference between both models being more significant in the case of the aggregated level similarity, showing the need for a similarity model specifically designed for aggregated levels, particularly in the case where concepts are related by part-of relations and only defined by their subordinated concepts. Indeed, in this case, the similarity of the MD model is reduced to the graph neighborhood similarity and lexical similarity, whereas with the similarity for aggregated levels (redefined model), similarity between features is implicitly considered in the recursive function. In addition, figure 4 shows that the curves have significant variations (as shown on graphs), reflecting the clustering of the similarity values between spatial zones. Distribution of the similarity values is not uniform but form clusters because variability of the features of spatial zones (in term of density, species, height, etc.) remains limited. Consequently, this results in high sensitivity of the recall and precision to weak variations of threshold during the test.

5 Conclusion and Perspectives

Our article presented a general approach of semantic mapping between different versions of geospatial data cube, which is different from other existing approaches in multidimensional structures evolution as it considers the case where relations between different structure versions are not identified a priori and are affected by semantic evolution. The suggested similarity model improves the precision of the semantic similarity assessment by evaluating similarity not only between concepts, but also between their features. The model is flexible and generic and can incorporate any type of feature (sound, image, various languages, etc.) for which it is possible to define a similarity measure. Then, it can be used in a variety of cases as it allows the integration of various types of data. Results showed that the effectiveness of the redefined model is higher than that of the Matching Distance model. More work is necessary in order to integrate these preliminary results to the SOLAP tool (Spatial OLAP), in order to improve quality of temporal queries results in evolving geospatial data cubes.

Acknowledgements. The authors wish to acknowledge the financial support of Canada NSERC Industrial Research Chair in Geospatial Databases for Decision Support.

References

1. Rivest, S., Bédard, Y., Proulx, M-J, Nadeau, M., Hubert, F., Pastor, J.: SOLAP Technology : Merging Business Intelligence with Geospatial Technology for Interactive Spatio-Temporal Exploration and Analysis of Data, ISPRS Journal of Photogrammetry and Remote Sensing 60: (2005) 17-33
2. Body, M., Miquel, M., Bédard, Y., Tchounikine, A.: A Multidimensional and Multiversion Structure for OLAP Applications. In Procceedings of the 5th ACM international workshop on Data Warehousing and OLAP (2002)
3. Thurnheer, A : Les Changements dans les Bases de Données Analytiques. In: Information als Erfolgsfactor, Teubner, Stuttgart (2000)
4. Eder, J., Koncilia, C., Mitsche, D.: Automatic Detection of Structural Changes in Data Warehouses. Lecture Notes in Computer Science Vol. 2737 (2003) 119-128
5. Mendelzon, A.O., Vaisman, A.A. (2000) Temporal Queries in OLAP. Proceedings of the 26th VLDB Conference, Cairo, Egypt
6. Eder, J., Koncilia, C., Morzy, T.: The COMET Metamodel for Temporal Data Warehouses. In Proc. of the 14th Int. Conference on Advanced Information Systems Engineering (CAISE'02), Toronto, Canada (2002)
7. Morzy, T., Wrembel, R.: On querying Versions of Multiversion Data Warehouse. DOLAP'04, Washington, DC, USA (2004).

8. Gruber, T.R.: Toward Principles for the Design of Ontologies used for Knowledge Sharing. International Journal of Human–Computer Studies,Vol. 43 (1995) 907–928
9. Klein, M.: Combining and Relating Ontologies: an Analysis of Problems and Solution. In IJCAI-2001 Workshop on Ontologies and Information Sharing, Seattle, WA (2001) 53-62
10. Eder, J., Koncilia, C.: Modelling Changes in Ontologies. OTM Workshops 2004, LNCS, 3292 (2004) 662-673.
11. Sure, Y., Staab, S., et Studer, R.. On-To-Knowledge Methodology (OTKM). In S. Staab et R. Studer (Eds.), Handbook on Ontologies, Springer Verlag (2004) 117-132
12. Ehrig, M., Sure, Y.: Ontology Mapping – An Integrated Approach. Proceeding of the First European Semantic Web Symposium, ESWS 2004 Heraklion, Crete, Greece (2004).
13. Su, X., Gulla, J.A.: Semantic Enrichment for Ontology Mapping. F.Meziane, E. Metais (Eds.) : NLDB 2004, LNCS 3136 (2004) 217-228
14. Euzenat, J., Valtchev, P.: Similarity-Based Ontology Alignment in OWL-Lite. In the 16th European Conference on Artificial Intelligence, Valencia, Spain (2004)
15. Rada, R., Mili, H., Bicknell, E., Blettner, M.: Developement and Application of a Metric on Semantic Nets. IEEE Transactions on Systems, Man and Cybernetics 19(1) (1989) 17-30
16. Wu, Z., Palmer, M.: Verb Semantic and Lexical Selection. In Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistic, Las Cruses, New Mexico (1994)
17. Jiang, J.J., Conrath, D.W.: Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy. In Proceedings of the International Conference on Research in Computational Linguistic, Taiwan (1998)
18. Hirst, G., St-Onge, D.: Lexical Chains as Representations of Context for the Detection and Correction of Malapropisms. In Christiane Fellbaum (editor), WordNet: An Electronic Lexical Database, Cambridge, MA: The MIT Press (1998)
19. Tversky, A.: Features of Similarity. Psychological Review 84(4) (1977) 327-352
20. Rodriguez, M.A.: Assessing Semantic Similarity Among Entity Classes. Thèse de doctorat, University of Maine (2000)
21. Zurita, V.: Semantic-Based Approach to Spatial Data Sources Integration. PhD Thesis, Universitat Politecnica de Catalunya (2004)
22. Resnick, P.: Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language. Journal of Artificial Intelligence Research 11 (1999) 95-130
23. Seco, N., Veale, T.,Hayes, J.: An Intrinsic Information Content Metric for Semantic Similarity in WordNet. In Proceedings of ECAI 2004, the 16th European Conference on Artificial Intelligence (2004)
24. Gärdenfors, P.: Conceptual Spaces: The Geometry of Thought.Cambridge, MA, MIT Press (2000)
25. Schwering, A.: Hybrid Model for Semantic Similarity Measurement. OTM Conferences (2) (2005) 1449-1465
26. Salton, G., Buckley, C.: Improving Retrieval Performance by Relevance Feed-Back. Journal of the American Society for Information Science, Volume 14, Issue 4 (1990) 288-297
27. Knappe, R., Bulskov, H., Andreasen, T.: On Similarity Measures for Content-Based Querying. In O. Kaynak, editor, Proceedings of the 10th International Fuzzy Systems Association World Congress (IFSA'03), Instanbul, Turkey (2003) 400–403
28. Carbo-Dorca, R., Besalu. E.: A General Survey of Molecular Quantum Similarity. J. Mol. Struct-Theochem 451(1-2) (1998) 11-23