Specifying Spatio Temporal Relations for Multimedia Ontologies

Karthik Thatipamula[1], Santanu Chaudhury[1], and Hiranmay Ghosh[2]

[1] Department of Electrical Engineering, IIT Delhi
kashyap_karthik@yahoo.com, santanuc@ee.iitd.ac.in
[2] Research Group, Tata Infotech Limited, New Delhi
hiranmay.ghosh@tatainfotech.com

Abstract. This paper present a novel framework for formal specification of spatio-temporal relations between media objects using fuzzy membership. We have illustrated its use in multimedia ontologies and have described a reasoning framework for creating media based descriptions of concepts.

1 Introduction

An ontology to support multimedia applications needs to deal with media properties of concepts. Traditional ontology languages, e.g. OWL, does not support explicit assignment of media properties with concepts, and reasoning with the intrinsic uncertainties involved in media based concept recognition. We have proposed an extension of OWL to overcome these limitations [1]. The extended language M-OWL enables formulation of media based description of concepts and concept recognition through observation of media properties. Spatial and temporal relations between media objects are important for characterizing media events. We describe a new scheme for formal description of such spatio-temporal relations in this paper. We also describe a reasoning scheme to construct a media-based description of the concepts using these relations.

Papadias et al [2] have proposed a scheme for formal specification of spatial and temporal relations. This scheme is based on the relative positioning of the end points of the events in space and time axes. However, these definitions are crisp and cannot satisfactorily express the intrinsically uncertain nature of nearness relations. Moreover, it characterizes an event with the minimum bounding rectangle with the projections of its end-points on the space and time axes and cannot cope up with the relations associated with concave events. We have extended this scheme with fuzzy membership functions and additional relations based on RCC model [3] to overcome these limitations. We have also introduced a reasoning scheme that enables construction of media property based descriptions for concepts with these spatio-temporal relations.

The rest of the paper is organized as follows. Section 2 describes the motivation behind extension of OWL for multimedia Ontologies. Section 3 describes a new way of specifying semantics to spatio temporal relations and the extension

S.K. Pal et al. (Eds.): PReMI 2005, LNCS 3776, pp. 527–532, 2005.

to M-OWL to incorporate this information. Section 4 describes a Bayesian network based inference scheme to reason in context of spatio temporal relations. Section 5 concludes this paper.

2 Multimedia Ontology and Its Representation

A concept manifests itself as some media patterns in a multimedia document. The observation of the expected media patterns is the key to concept recognition. The traditional ontology languages, e.g. OWL, does not have a formal semantics for associating media patterns with the concepts. As a result, it is not possible to do reasoning with the media properties, e.g. a monument made of marble should "inherit" its color and texture properties. The crisp Description Logic based reasoning with OWL does not support the inherent uncertainties with semantic interpretation of media data.

Multimedia ontologies should enable construction of a model for the concepts in terms of media objects and their relations and enable reasoning with these properties and associated uncertainties. We have proposed M-OWL as an extension of OWL to incorporate these features [1]. M-OWL enables construction of an Observation Model (OM), which is a media-based description of a concept. It is organized as a Bayesian Network with the root node representing the concept and the leaf nodes representing a set of media patterns that are expected in a multimedia document where the concept materializes. The causal links connecting the concept with the media patterns in the Bayesian Network represents the uncertainties that are associated with semantic interpretation of media data. Observation of the expected media patterns in a multimedia document leads to concept recognition through belief propagation in the Bayesian Network [5].

Spatial and temporal relations between the media objects is another important aspect of multimedia. It should be possible for a multimedia ontology to Provide the capability for formal definitions for these relations and to reason with them. For example, a "goalScore" event in a football game can be described as ((`ball` *inside* `goalPost`) *followedBy* `cheering`). It should be possible to define the semantics of the relations *inside* and *followedBy* formally in a multimedia ontology, and use them to construct an observation model for the "goalScore" event with the media properties of the constituent concepts, namely "ball", "goalPost" and "cheering". Moreover, it should be possible to specify the degree of memberships for the spatio-temporal relations for different configurations of the connected concepts. We introduce a new method for formally specifying the spatio-temporal relations and reasoning with them in subsequent sections of this paper.

3 Encoding Spatio-temporal Relations

Padadias et al [2] proposes formal encoding of relation between two events in space-time encoded as a set of binary strings. If [a,b] be a closed and continuous 1D interval, we can identify five distinct regions of interest: $(-\infty, a)$, $[a, a]$, (a, b),

$[b, b]$, $(b, +\infty)$. The relationship between this primary interval and any other interval [z,y] can be specified by a binary string $< t, u, v, w, x >$ representing empty or non-empty intersections of [z,y] with the five regions of interest for [a,b]. Thus, the relation between two media events in a multimedia document can be represented by a 3-tuple $< C\_t, X\_p, Y\_p >$ denoting the binary intersection-strings in time, X and Y axes respectively.

This encoding scheme assumes that the events are are represented by convex regions in space-time and fails to unambiguously encode the relations when any of the events is concave. For example, the region B is contained in the region A in figure 1(b), while it is not in figure 1(a). Papadias'e encoding scheme cannot distinguish between the two scenarios. Motivated by RCC [3], we have introduced one more string C_s, which specifies the overlap or disconnectedness between the regions to solve this problem. C_s is represented by a string $< f\_0, f\_1, f\_2 >$ where

- f_0 - $A - (A \bigcap B)$
- f_1 - $B - (A \bigcap B)$
- f_2 - $(A \bigcap B) - (A \bigcap^* B)$ where $\bigcap^*$ denotes the regularized set operation.

Thus, the spatio-temporal relation is unambiguously defined as the 4-tuple $<C\_t$, X_p, Y_p, $C\_s >$.

Another weakness in Papadias'es scheme is in the definition of nearness relation in terms of a constant δ. The regions (a-δ,a) and (b,b+δ) are considered to be the neighborhood of [a,b]. This approach is unsatisfactory, since the definition of the constant δ is arbitrary and the nearness should not have a sharp cutoff and a-δ or b+δ. We solve this problem by introducing fuzzy membership functions in place of the binary intersection variables.

The membership function models the relation between intervals in a soft parametric fashion. For example consider the encoding of the fourth variable w for the temporal relation *followedBy* as shown in figure 2. The three different events $[a, b]$, $[a1, b1]$ and $[a2, b2]$ are all after the event $[x, y]$ but due to the membership function defined, the variable will have a very high value for $[a1, b1]$ qualifying it for the relation *followedBy*, where as the event $[a3, b3]$ will have a very low value and will not qualify.

Therefore in general, we can define temporal relations in 1D to be a 5-tuple $C_t =< t, u, v, w, x >$ each of which represents the membership function of the regions of intersection. The scheme of Papadias [2] is a special case of our scheme.

To encode the above primitives in M-OWL we propose a *STOp* class (Spatio Temporal Operator) which can have one or more of the components $< hasCs >, < hasXp >, < hasYp >, < hasCt >$. Each of them is a datatype property having the range as a string.The string gives the values of the parameters C_s, X_p, Y_p and C_t for any given spatio-temporal relation. Each component has a again five values $< t >, < u >, < v >, < w >$ and $< x >$, where each component specifies the membership function associated with each variable. The membership functions can be specified from a library of predefined functions and the parameters of the function can be specified in the definition. The function in figure 2 can be represented as

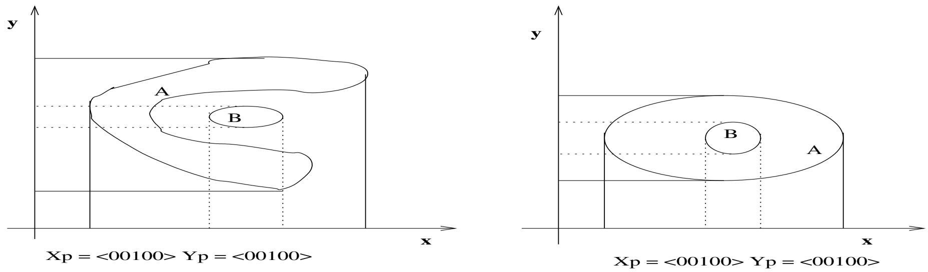

Fig. 1. Ambiguity due to concave regions

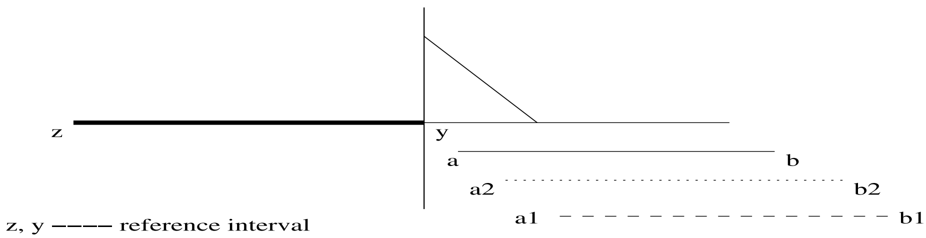

Fig. 2. Encoding a fuzzy temporal relation. The membership function is also shown.

a Piece-wise linear model as $<(-\infty\ 0),(0\ 0),(\ 0\ 1),(0.5\ 0.5),(1\ 0),(+\infty\ 0)>$. Not all of the components are required depending upon the relation as purely spatial or purely temporal or in general spatio-temporal. It is not necessary to specify all the membership functions for every relation.

4 Reasoning for Description Generation of Abstract Concepts

We define two more attributes for relations "propagate" and "hierarchy" as follows:

- **Hierarchy:** A relation between two concepts A and B has an attribute hierarchy true if an example of B is an example of A i.e B is a subconcept of A.
- **Propagate:** A relation between two concepts A and B has propagate true if the media features of A are inherited by B independent of the concept hierarchy.

With the above definitions, we propose a solution of the unique problem of Concept Description Generation.This involves the construction of media based description of concepts,via propagation and inheritance of media examples in an Observation Model. The steps for construction of OM from a M-OWL encoding and generating a description are

1. Every concept “C” is mapped to a node in the OM.
2. Every Property “P” with hierarchy true and with domain “D” and range “R”is mapped to a node in the translated network which has the set of individuals in the domain who have this property “P” .
3. Every property node has one more node as its child which has the individuals of the global range of the property.
4. Every spatio temporal relation is translated to a intermediate node between the composite concept and the sub concepts. It has a child node which denotes the operator defined for that property. The above rules construct an OM from any M-OWL encoding.
5. Now to generate a description of the composite concept we start with the media examples of the leaf nodes and transfer them *upwards* until we get a spatio-temporal node.
6. At a spatio temporal node the examples of individual concepts are composed according to the relation and then the composed example is transferred upwards.
7. The above step is repeated for each spatio temporal node until we reach the composite concept.

For the example concept of “goalScore” in section 2 the M-OWL encoding given in appendix 5. Following the rules mentioned above we get the OM shown in figure 3.The “STOp” nodes represent the operator of the spatio temporal relation. Note that the operator does not tell us any specific positioning of the examples of “goal” and “goalPost”. As long as the positions are in accordance of the operator this will be an example of “goalScore” With this we can construct many such examples of the concept “goalScore”. Similar scheme can be used for construction OM of an abstract concept in terms of media properties.

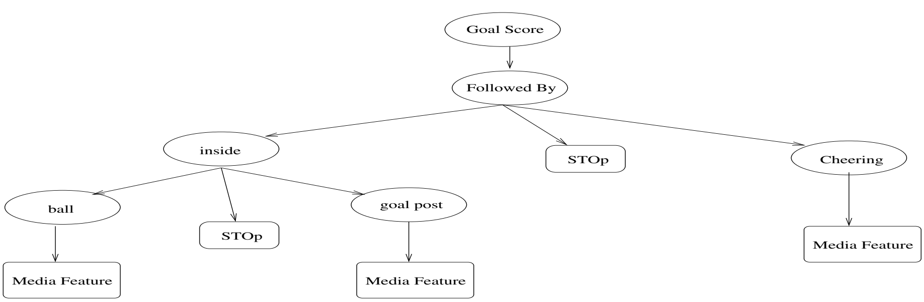

Fig. 3. Observation Model of spatio temporal relations

For the problem of concept recognition and retrieval we construct the propagate graph with all propagate relations and an OM is constructed as per the rules mentioned above.The conditional probabilities specified between concepts are included and the OM is now a belief network. A belief network based retrieval application with media features has been discussed in [4].

5 Conclusion

The above scheme of representing Spatio Temporal Relations and combining them with propagate and hierarchy gives a flexible way to deal with problems like concept description and concept recognition in in Multimedia Ontologies. We are working on other issues like – what is the interaction between an OWL ontology and a multimedia ontology and to what extent they can work together and to what extent Description Logic based inferencing of OWL can be applied to a multimedia ontology.

Acknowledgements. This work has been supported by the grant for DST project "E-Video: Video Information Processing with Enhanced Functionality."

References

1. H. Ghosh, S. Chaudhury, K. Kashyap, and B. Maiti, *Ontologies in Context of Information Systems*, ch. Ontology Specification and Integration for Multimedia Applications. Kluwer Press (Accepted for Publication).
2. Papadias, "Approximate spatio temporal retrieval," *ACM Transactions on Information Systems*, vol. 19, pp. 53–96, January 2001.
3. Cohn, "Qualitative spatial representation and reasoning with the region connected calculus," *GeoInformatica*, vol. 1, pp. 1–44, 1997.
4. S. Chaudhury and H. Ghosh, "Distributed and reactive query planning in RMAGIC:An agent based multimedia retrieval system," *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, pp. 1082–1095, September 2004.

Appendix: M-OWL Encoding for Goalscore

```
<mowl:STOp rdf:ID="followedBy">
 <hasCt> <w>-INF 0 0 1 0.5 0.5 1 0 INF 0</w> </hasCt>
</mowl:STOp>
<mowl:STOP rdf:ID="inside">
<hasXp><v>-INF 0 0 0 0.5 0.5 1 1 3 1 3.5 0.5 4 0 INF 0</v></hasXp>
<hasYp><v>-INF 0 0 0 0.5 0.5 1 1 3 1 3.5 0.5 4 0 INF 0</v></hasYp>
</mowl:STOp>
<mowl:MediaFeature rdf:ID="ballShape">
  <Mpeg7: ..  ... a shape descriptor for ball shape </Mpeg7:..
</mowl:MediaFeature>
<mowl:MediaFeature rdf:ID="goalPostShape">
  <Mpeg7: ..  ... a shape descriptor for goal post shape </Mpeg7:..
</mowl:MediaFeature>
<mowl:MediaFeature rdf:ID="audioPatternForCheering">
  <Mpeg7: ..  ... an audio descriptor for cheering </Mpeg7:..
</mowl:MediaFeature>
<owl:Class rdf:ID="ball">
  <hasMediaFeature>ballShape</hasMediaFeature>
</owl:Class>
<owl:Class rdf:ID="goalPost">
  <hasMediaFeature>goalPostShape</hasMediaFeature>
</owl:Class>
<owl:Class rdf:ID="cheering">
  <hasMediaFeature>audioPatternForCheering</hasMediaFeature>
</owl:Class>
<owl:Class rdf:ID="goalScore">
  <mowl:followedBy rdf:parseType="Collection">
    <OWL:Class rdf:about="cheering"/>
    <mowl:inside rdf:parseType="Collection">
      <owl:Class rdf:about="goalPost"/>
      <owl:Class rdf:about="ball"/>
    </mowl:inside>
 </mowl:followedBy>
</owl:Class>
```