Multi-dimensional knowledge representation with a fuzzy extension

Debasis Mitra
Department of Computer Science, Jackson State University
P.O. Box 18839
Jackson, MS 39217
Tel: 601-968-8252 Fax: 601-968-2478
E-mail: dmitra@stallion.jsums.edu

Abstract

This paper presents some preliminary results of our current attempts to develop a hybrid multi-dimensional knowledge representation scheme which can handle both the incompleteness and uncertainty. We have started from the quantitative temporal constraint information, and extended it with possibilistic quantifiers. We have also extended the former towards a multi-dimensional constraint-based formalism. Finally we have combined these two extensions under a multi-dimensional possibilistic scheme. The context of spatio-temporal reasoning is one of the motivations for developing such a hybrid scheme. However, the major motivation for our work is coming from the muti-dimensional data modeling research[SHJM96] within the database area.

Identifying notions: Quantitative temporal constraint propagation, Fuzzy constraint propagation, Spatio-temporal knowledge representation, Multi-dimensional data modeling.

1 Introduction

There have been some attempts to enhance the incompleteness of constraints on temporal knowledge representation[MSGH94]and spatial knowledge representation[GHP94] schemes with uncertainty management capabilities. Vila et al[VG94] have addressed the same issue in the context of quantitative knowledge rather than qualitative knowledge. In this paper we have attacked the same problem as handled by Vila et al, but we have proposed a different fuzzy composition operator (section 3) than a rather intuitive one suggested by them. The reason behind this choice is explained in this article.

We have also extended the quantitative temporal knowledge representation scheme towards multi-dimensional knowledge representation (MDKR) in the section 4, whereas the fuzzy extension of MDKR is proposed in the section 5. The demand for a MDKR scheme comes from varieties of fields including AI. While within AI, spatial and temporal knowledge representation is related mostly to the common sense reasoning, data modeling researchers within the database community recently are groping with the problem of handling multi-dimensional data[CCS93]. In order to extend the capabilities of multi-dimensional data models, an MDKR is necessary. These issues are briefly discussed in the section 6.

Section two of this paper lays the foundation of our work with a revisit of the one-dimensional (temporal) knowledge representation scheme proposed by Dechter et al[DMP91] which have been extended by Vila et al[VG94] for handling uncertainty. The last section concludes this article.

2 Background

Quantitative temporal constraint network (QTCN) is a graph whose each node represents the timing of an event, and whose directed arcs represent time-intervals between pairs of such events, arcs being pointing towards the future. The label on an arc represents the possible range of the interval between the end points. For example, A –[10, 20]–> B indicates B happened after A by 10 to 20 units of time. The range here implies incompleteness of knowledge. When the knowledge is complete the range would be a pair of the same value, e.g., [10,10]. Dechter et al and Meiri[DMP91,Mei96] have extensively studied the constraint propagation in this framework. A network with only convex intervals (which do not have holes in them, e.g., {[10,15][17,20]}) as labels are called *simple* network, on which the problem of constraint propagation is polynomial, otherwise the problem is NP-complete. In this paper we will focus only on simple networks.

A temporal constraint propagation (TCP) algorithm is centered around four basic operations between the label-type of temporal constraint networks. These operations are the set union, the set intersection, inverting and the composition. For QTCNs labels are the intervals of time for which the set union and the set intersection are trivial operations. The inverting operation changes the direction of an arc, and for a label [x, y] the new label becomes [-y, -x] on the reverse arc. It is the binary composition operation which is unique for the temporal constraint propagation problem, and the focus of our fuzzification approach presented in this paper. The composition of the two convex labels
A –[x, y]–> B and B –[a, b]–> C is A –[x+a, y+b]–> C.

What happens with a composition operation could be better understood with the following three steps. (1) If A –[x, x]–> B and B –[a, a]–> C (complete constraints, not ranges), then A –[x+a, x+a]–> C. (2) When A –[x, x]–> B and B –[a, b]–> C (one constraint is complete, the other one is a range), then A –[x+a, x+b]–> C. This is actually obtained by taking composition of each pair of points $(\forall i(a \leq i \leq b)(x, i))$ between the two labels as in the step 1, and then taking a union of their results. (3) Thus, when the composition operation is performed between two incomplete constraints (as mentioned in the last paragraph) the same is also done between each pair of points in the operand labels, and then finally a union operation is done to obtain the resulting label on the arc from A to C. This observation regarding how the composition operation is actually done affects our choice of the fuzzy composition operator. Incidentally, this is exactly the same process which is applied for the composition operation between disjunctive qualitative temporal relations for a qualitative interval constraint network[All83].

Constraint propagation on QTCN involves tightening the ranges on labels further than what has been given, in order to create a *path consistent* network. Dechter et al[DMP91] have suggested two ways of propagating constraints on a QTCN. The first one is based on splitting each arc
A –[x, y]–> B into two opposite arcs A –[x]–> B and B –[-y]–> A and then running Floyd's shortest path algorithm on it. The second approach is to run a *path consistency* algorithm by using the composition operation as described above. The algorithm runs until the network attains a stability without further tightening any more labels, or until an inconsistency is detected as a *Null*

label on some arc.

3 Fuzzy temporal knowledge representation

Fuzzy temporal knowledge representation schemes extend the incompleteness of constraints (e.g., range on the labels of a QTCN) with possibilistic uncertainty values (see [DP90] for more details on possibility theories). Thus in a fuzzy QTCN each label would be associated with a distribution function over the range as the fuzzy quantifier for that label. In the classical case, i.e., when no such distribution is provided, a default fuzzy distribution is assumed which takes a rectangular shape over the range associated with a convex label. Such a rectangular distribution of the fuzzy quantifier has the value one over the range and the value zero outside the range.

In order to develop a fuzzy TCP algorithm one has to first extend the four operations mentioned in the last section with the corresponding fuzzy operations for the quantifiers. The canonical fuzzy operators for the set union and the set intersection operations are *max* and *min* respectively. The inverse operation would correspond to an operation inverting the domain of a function, i.e., a fuzzy quantifier $\mu(x)$ will become $\mu(-x)$.

Before we define the fuzzy composition operator, a point has to be noted over the classical composition operation described in the last section. Assume, $A - x -> B$ and $B - y -> C$, where x and y are two *discretized* ranges [1,2,3, ..., n] and [1,2,3, ..., n]. Assume their composition is $B - z -> C$, where z is the range [2,3,4, ..., 2n]. Each point in z is constituted from some number of pairs of points in x and y. Thus, the first point 2 in z is constructed from the pair (1, 1) in the operands of the composition. We call this number of pairs of supporting pairs as a **support factor** corresponding to each point in z. This factor grows towards the middle of z and then it falls again towards the end following a bell-shaped distribution. Maximally supported points on z are on its middle around n. For example, the point n is constituted by n-1 number of pairs {(1, n-1), (2, n-2), ..., (n-2, 2), (n-1, 1)} in x and y (**observation 1**). The distribution of this *support* (number of pairs constituting the resulting point) is always symmetric (**observation 2**). Even in the qualitative interval-case, some primitive relations in the results of a disjunctive composition operation is better supported than others by a higher number of pairs of primitive composition operations (refer to [All83] for details of the composition operation in that context).

The observation (1) in the above paragraph leads us to believe that even in a classical case (with the default fuzzy distribution), the middle part of the resulting range from a composition operation should have a higher possibility value than those towards ends, because of the fact that the middle part is better *supported* by the operands (**principle 1**). In order to conform to this principle we suggest the following operator for constructing the fuzzy distribution over the resulting range from a composition operation.

Step 1. If A –$[(x, x, \mu_1(x))]$–> B (complete but uncertain constraint) and B –$[(a, a, \mu_2(a))]$–> C, then A –$[(x+a, x+a, min(\mu_1(x), \mu_2(a)))]$–> C. This conforms to the classical case when one of the two fuzzy operand quantifiers is zero indicating an impossibility of the corresponding time-value (x or a) on the relevant arc resulting in a zero possibility of the resulting label ($x+a$).

Step 2. Suppose, A –$[(x, x, \mu(x))]$–> B and B –$[a, b, F]$–> C, where $F(i)$ is a fuzzy distribution over the range $a \leq i \leq b$. Then A –$[x+a, x+b, G]$–> C, where $G(i) = min(\mu(x), F(i)), \forall i (a \leq i \leq b)$.

Step 3. When A –$[(x, y, F_1)]$–> B and B –$[a, b, F_2]$–> C,
then A –$[x+a, y+b, G]$–> C, where

$$\forall k(x+a \leq k \leq y+b)[G(k) = \frac{\sum_{i=x}^{y}\sum_{j=a}^{b}\{min(\mu_{F_1}(i), \mu_{F_2}(j)) \mid k = i+j\}}{min(y-x, b-a)}]$$

The summation takes care of the higher support in the middle of the resulting distribution. The denominator is to normalize the resulting fuzzy value after the summation so that it remains within [0,1]. It is easy to verify that the number of supports is bounded by the minimum of the two ranges (**observation 3**). Vila et al[VG94] have suggested a *max of min* formula for this operation. But as we have discussed before, that will not take care of the *support* factor.

In practice, the operand ranges have to be discretized to do any fuzzy composition operation. If each of the operands has n number of discretized points, then a fuzzy composition operation will clearly have an $O(n^3)$ complexity (coming from operations over $\forall k \forall i \forall j$). For a practical purpose, this could be brought down to a constant factor, if the *granularity* of time is controlled in order to generate discrete points with a pre-fixed upper bound on their number.[1]

4 Multi-dimensional knowledge representation

In a two-dimensional constraint network a label is a pair of two dimensional points indicating a rectangular range, as opposed to the scalar range in an one-dimensional network like the temporal network discussed in the previous sections. Thus, a 2-D constraint looks like
A -[$(x_l, y_l), (x_h, y_h)$]-> B, where suffices l and h indicate low and high values of the range. Here a range is a rectangle in the two dimensional plane. A range is discretized over a set of grid points depending on the granularity of the relevant dimensions. An n-dimensional constraint network will have pairs of n-dimensional points as its labels. A spatio-temporal constraint network could be considered as an example of such a network, (at most 4-D, with three dimensional spatial constraints and one dimensional temporal constraints between nodes, which are events located in space). For example, *accident1* -[(5m, 10s), (8m, 15s)]-> *accident2*, indicates that the two accidents are separated by 5 to 8 meters (on the road) and they happened with 10 to 15 seconds of interval.

For the purpose of constraint propagation one has to define the four operations mentioned in the section two. In a multi-dimensional set up, the set union, the set intersection and the inverse operations are trivial as in the one dimensional case. The composition operation will follow our understanding for 1-D case described in the section two. For a 2-D constraint network:
Step 1. If A -[(x_1, x_2), (x_1, x_2)]-> B and B -[(a_1, a_2), (a_1, a_2)]-> C (complete constraints), then A -[$(x_1 + a_1, x_2 + a_2)$]-> C, where a suffix indicates the corresponding dimension.
Step 2. When A -[(x_1, x_2), (x_1, x_2)]-> B and B -[$(a_1, a_2), (b_1, b_2)$]-> C, then A -[$(x_1 + a_1, x_2 + a_2), (x_1 + b_1, x_2 + b_2)$]-> C. This is just shifting the rectangle (second operand) by a vector (first operand).
Step 3. When A -[$(x_1, x_2), (y_1, y_2)$]-> B and B -[$(a_1, a_2), (b_1, b_2)$]-> C, then
A -[$(x_1 + a_1, x_2 + a_2), (y_1 + b_1, y_2 + b_2)$]-> C. As in the 1-D case, the point to point operation for

[1] In the qualitative case a natural upper bound exists which is 13 for the thirteen primitive relations. This is the limit of the cardinality of a label.

obtaining the resulting rectangle (observations 1 and 2, and the principle 1) has to be kept in mind here also.

Constraint propagation for a multi-dimensional constraint network is basically running a path-consistency algorithm independently on each of the dimensions. An inconsistency along any dimension implies inconsistency of the network.

5 Fuzzy multi-dimensional knowledge representation

In a multi-dimensional framework a fuzzy quantifier is a multi-dimensional function, e.g. for 2-D it is $\mu(x, y)$. As usual the set union, the set intersection, and the inverting operations are trivial. The composition operation is a combination of the same operation for the 1-D fuzzy case and that for the multi-dimensional case. The composition operation in this case could be considered as a generalized version of the two cases considered in the previous two sections. We are providing here the relevant formula for only the general (incomplete and uncertain) 2-D case, or as in the step 3 of our previous discussions.

Step 3. When A –$[(x_1, x_2), (y_1, y_2), F_1(x, y)]$–> B and B –$[(a_1, a_2), (b_1, b_2), F_2(x, y)]$–> C, then A –$[(x_1 + a_1, x_2 + a_2), (y_1 + b_1, y_2 + b_2), G(x, y)]$–> C, where

$$\forall k_x(x_1 + a_1 \leq k_x \leq x_2 + b_2)\forall k_y(y_1 + a_1 \leq k_y \leq y_2 + b_2)$$

$[G(k_x, k_y) =$

$$\frac{\sum_{i_x=x_1}^{x_2} \sum_{j_x=a_1}^{a_2} \sum_{i_y=y_1}^{y_2} \sum_{j_y=b_1}^{b_2} \{min(\mu_{F_1}(i_x, i_y), \mu_{F_2}(j_x, j_y)) \mid k_x = i_x + j_x, k_y = i_y + j_y\}}{min(x_2 - x_1, a_2 - a_1).min(y_2 - y_1, b_2 - b_1)}$$

]

Constraint propagation in this case is no longer just running a path-consistency algorithm independently on each of the dimensions. This is because the fuzzy quantifiers are not independent along different dimensions, rather they are multi-dimensional functions. A constraint propagation algorithm here is an extension of a path-consistency algorithm for 1-D networks. We are currently developing such an algorithm.

6 Usefulness of the MDKR

Both the spatial knowledge representation as well as the temporal knowledge representation communities are realizing for sometime that it is necessary to combine these two fields of investigation under one scheme of spatio-temporal knowledge representation. This is not only because there are enough commonalties of approaches between these two areas, but also because in many problem domains time and space appear in a conjugal fashion.

There are other situations when it is not just space and time, but multiple (even abstract or non-physical) dimensions appear together. Multi-dimensional database management is such a situation. On-line analytical processing (OLAP)[CCS93] is a front-end for a data warehouse, where data is primarily viewed as multi-dimensional for the purpose of analysis to be performed by business managers. Scientific and statistical data sets are inherently multi-dimensional[Sho97]. A constraint database targeted towards these areas will need not only a multi-dimensional data model, but also a multi-dimensional knowledge representation scheme, such as the one presented in this paper, if it wants to handle incomplete information. This kind of capability of handling incomplete information in a database would allow a user to enter constraints between pairs of data points rather than the exact coordinate of each of them. Addressing incomplete and fuzzy information is already being attempted in the industry, e.g. the Level-5 Quest(TM) of the Level-5 corporation (URL address: www.l5r.com).

The fuzzy-extension of a MDKR powers the latter with the capability of more expressiveness. Apart from that practical issue, studying a fuzzy extension often reveals some intricacies of the relevant KR scheme which would not be apparent otherwise. For example, our fuzzification attempt here has revealed that the supports for some elements in the result of a composition operation are more than those for others (refer the observation 1). A fuzzy-enhanced MDKR will also add power to the fuzzy database products like Level-5 Quest(TM), and to the current works on multi-dimensional data modeling[Sho97].

7 Conclusion

One of the major contributions in this paper is making the foundation of constraint propagation in quantitative temporal constraint networks more explicit. We have revisited the composition operation of the incomplete constraints. This understanding has allowed us to define firstly a new fuzzy composition operator, and secondly an extension from an 1-D temporal scenario to a multi-dimensional scenario where the dimensions could be space, time or any other abstract ones. The motivation for such MDKR comes mainly from the database area. However, our work could also be viewed as an attempt to generalize spatio-temporal knowledge representation schemes.

In this article we have also integrated the fuzzy temporal knowledge representation scheme with the multi-dimensional knowledge representation scheme in a seamless way. Our methodology here is also inherently compatible with the qualitative knowledge representation scheme. Our goal in this research is to develop a uniform technique for handling an integrated knowledge related to space, time, or any other dimension, where the knowledge could be quantitative or qualitative, and it could also be incomplete and uncertain. Developing a declarative language for expressing such knowledge, and developing algorithms for constraint propagation in the proposed hybrid scheme are natural future directions of our work.

Acknowledgment: I have had some productive discussions with Chen Liang during this work.

References

[All83] J. F. Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843, 1983.

[CCS93] E.F. Codd, S.B. Codd, and C.T. Salley. Providing OLAP (online analytical processing) to user-analysts: an it mandate. *E.F. Codd Associates, URL: http://www.arborsoft.com/papers/coddTOC.html*, 1993.

[DMP91] R. Dechter, Itay Meiri, and J. Pearl. Temporal constraint networks. *Artificial Intelligence*, 49:61–95, 1991.

[DP90] Didier Dubois and Henri Prade. An introduction to possibilistic and fuzzy logics. In Glenn Shafer and Judea Perl, editors, *Readings in Uncertainty Reasoning*, Morgan Kafman, 1990.

[GHP94] Hans Guesgen, Joachim Hertzberg, and Anne Philpott. Towards implementing fuzzy Allen relations. In *Proceedings of European Conference on AI (ECAI)*, 1994.

[Mei96] Itay Meiri. Combining qualitative and quantitative constraints in temporal reasoning. *Artificial Intelligence*, 87(2):343–385, 1996.

[MSGH94] Debasis Mitra, Padmini Srnivasan, Mark L. Gerard, and Adrian E. Hands. A probabilistic interval constraint problem: fuzzy temporal reasoning. In *World Congress on Computational Intelligence (FUZZ-IEEE'94)*, 1994.

[SHJM96] Arie Shoshani, Preston Holland, Janet Jacobsen, and Debasis Mitra. Characterization of temporal sequences in geophysical databases. In *Proceedings of Scientific and Statistical Databases conference*, 1996.

[Sho97] Arie Shoshani. OLAP and statistical databases: similarities and differences: a tutorial. In *Proceedings of PODS (to appear)*, 1997.

[VG94] Lluis Vila and Lluis Godo. On fuzzy temporal constraint networks. *Mathware and Soft Computing*, 3:315–333, 1994.