Spatial Data Integration with Qualitative Integrity Constraints

J. O. Wallgrün[1], F. Dylla[2]

[1]IRTG GRK 1492 Semantic Integration of Geospatial Information
[2]SFB/TR8 Spatial Cognition
University of Bremen, Enrique-Schmidt-Str. 5, 28359 Bremen, Germany
Email: wallgruen|dylla@informatik.uni-bremen.de

1. Introduction

Methods for automatically integrating spatial information from different sources are becoming more and more important. In many integration scenarios, the result is supposed not only to be consistent but also satisfy a set of integrity constraints (IC). Often these constraints can be captured by the qualitative spatial relations from a qualitative spatial calculus (QSC) developed in the area of qualitative spatial reasoning (QSR) (Cohn and Renz 2007). For instance the requirement that "every city district has to be spatially *contained in* a city" can be modeled using the RCC8 (Randell et al. 1992) or 9-intersection calculus (Egenhofer 1989). In this work, we propose a framework to approach the spatial integration problem under such kind of qualitative IC rules, allowing the input data to be quantitative (e.g., polygonal), qualitative, or mixed. Our approach first transfers all information to the qualitative level, then uses an approach for merging qualitative spatial information including the resolution of conflicts, and finally adapts the quantitative input data to be in accordance with the merging results. We provide an overview on this framework, focusing on our method for merging qualitative information under ICs.

2. General Framework

Our general integration approach is illustrated in Figure 1. The input consists of several knowledge bases (KB). Given a set of ICs making statements about qualitative spatial relations that have to hold between certain kinds of objects, the strategy is to first compute a suitable merging result on the qualitative level before merging and adapting the quantitative data. Our framework consists of four core modules that will be described below: (1) a qualification module, (2) a module to interpret ICs, (3) the qualitative merging module, and (4) a module for adjusting quantitative data.

To illustrate our approach, we will use the simple example of merging two quantitative KBs containing (rather incomplete) information about an imaginary city C and three of its districts F, G, and H. KB_1 contains geometric information about F, G, and H given in the form of polygons $P_{F,1}$, $P_{G,1}$, and $P_{H,1}$ (see Figure 2a; for simplicity we only use rectangles in the example). KB_2 contains the same kind of information for C ($P_{C,2}$) and also for H ($P_{H,2}$). We further assume that the ICs are: (1) If y is a district of city z then y is a non-tangential or tangential proper part of z (RCC8 relations NTPP or TPP), and (2) two city districts do not intersect (i.e., their relation is disconnected (DC) or externally connected (EC)). Moreover, we assume that we have additional semantic information coming along with the KBs that tells us that C is a city and F, G, and H are all city districts of C. As Figure 2a shows there are several contradictions and violations of the ICs in the KBs (e.g., $P_{F,1}$ overlapping both $P_{C,2}$ and $P_{H,2}$).

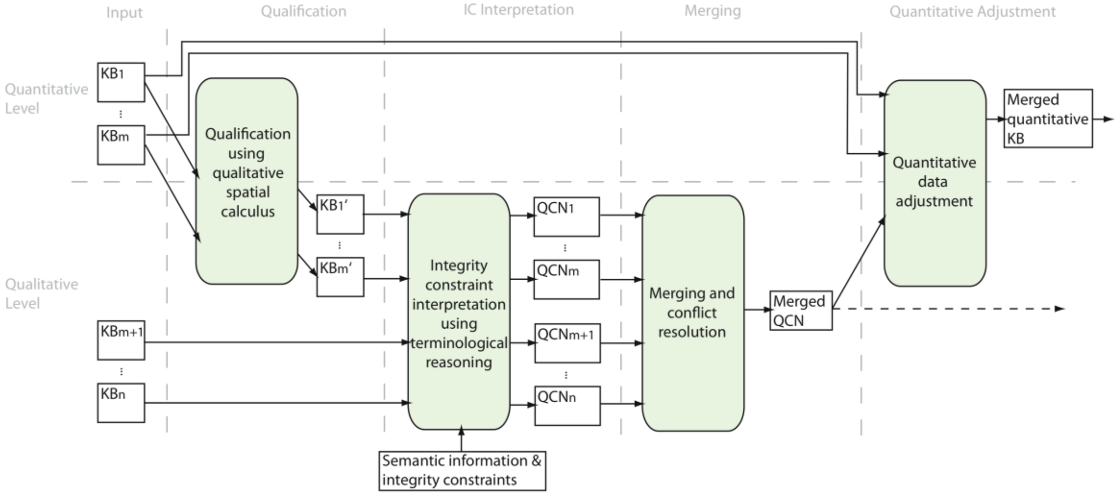

Figure 1. Our spatial data integration framework.

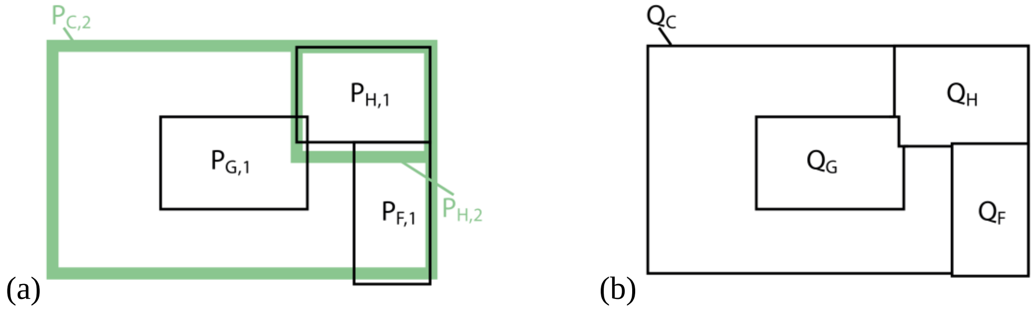

Figure 2. (a) Polygons of KB$_1$ and KB$_2$ and (b) the final integration result.

2.1 Qualification of Input Data

We refer to the process of transferring information from quantitative KBs to the qualitative level as *qualification*. In our example, the qualification with RCC8 would result in KB$_1$' containing the relational statements F {DC} G, G {O} H, and F {EC} H, while KB$_2$' consists of C {TPP} H. Computing qualitative relations is mainly an application of computational geometry methods which for RCC8 is already supported by most GIS systems. In addition, our own QSR toolbox SparQ (Wallgrün et al. 2007) supports qualification for a large number of QSCs.

2.2 Integrity Constraint Interpretation

In this work, we consider ICs that, when applied, yield concrete qualitative spatial relations that have to hold between certain tuples of objects from the input KBs. Interpretation of the qualitative ICs in the context of the semantic background information means to derive these spatial relation tuples and may involve complex terminological reasoning to determine the applicability of an IC for a given pair of objects (e.g., performed by a DL reasoner). In our example, the involved reasoning is trivial and the result is that the RCC8 relations between F, G, and H all have to be EC or DC, while those of F, G, H to C have to be NTPP or TPP. The output is a special sort of qualitative constraint network (QCN) for every KB that contains the qualitative relations from the input information as well as the restrictions derived from the ICs (in the following referred to as R$_{IC}$). Figures 3a and 3b show the resulting QCNs for our example with the derived restrictions given below the lines in the labels (U stands for the union of all RCC8 base relations).

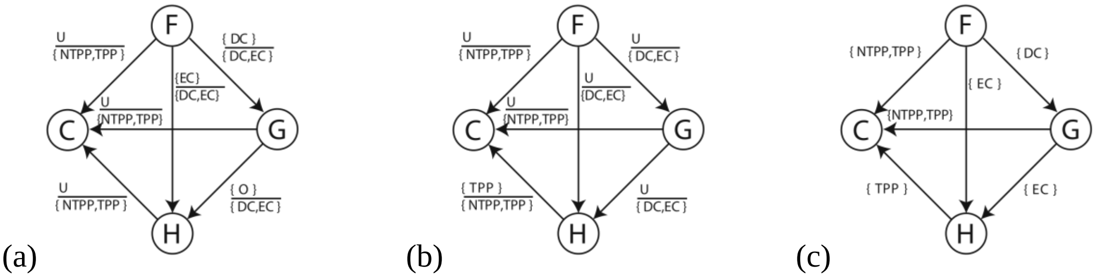

Figure 3. (a,b) QCNs for KB_1 and KB_2 and (c) the merging result.

2.3 Merging and Conflict Resolution

The goal of merging is to derive a single consistent QCN from a set of potentially conflicting input QCNs. Moreover, this QCN has to satisfy the ICs, meaning all constraints need to consist of base relations from the restrictions R_{IC}. In our approach we employ our own method developed for merging qualitative spatial databases (Wallgrün and Dylla 2010) and extend it to take into account the restrictions stemming from the integrity constraints. The theory behind this approach is further explained in Section 3. In our example, the approach yields the network shown in Figure 3c. One result is that the relation between G and H has been determined to be EC.

2.4 Adjustment of Quantitative Data

Given the QCN from Figure 3c, the final step is to combine and adapt the quantitative data so that it satisfies the qualitative relations. Developing general solutions to this problem able to deal with arbitrary QSCs is an open research problem. One approach could be to consider the problem as a constrained optimization problem. Currently, we employ an approach tailored for RCC8 relations and only able to correct certain deviations. To correct an overlap relation to EC, intersecting areas between the involved objects are basically cut into half. Adapting our input data to fit the QCN from Figure 3c results in the new quantitative KB illustrated in Figure 2b. Disjunctions in the QCN such as F {TPP, NTPP} C have been resolved by taking the relation that requires smaller adaptations on the quantitative level (in this case TPP).

3. Qualitative Merging with Integrity Constraints

Approaches for merging multiple QCNs into a single QCN and resolving conflicts have been developed over the recent years (Dylla and Wallgrün 2007, Condotta et al. 2008, Wallgrün and Dylla 2010). The idea is to find a consistent network that is closest to the input QCNs. The required distance measure is derived by first using the shortest path distance in the conceptual neighborhood graph (Freksa 1991) to describe the distance between base relations of the used QSC. Using two aggregation operators (e.g., sum or max), this distance is then first extended to describe the similarity of two scenarios, that is between QCNs in which all constraints consist of a single base relation, and finally to define the distance $d(s,\mathcal{N})$ between a scenario *s* and the entire set $\mathcal{N}$ of input QCNs. Given this distance function, we defined merging operators that yield the union of all scenarios that are at most as distant to $\mathcal{N}$ as the closest consistent scenario. We now adapt this definition by demanding that the scenarios also have to satisfy the ICs. To do so, we use the notation $\langle\!\langle\cdot\rangle\!\rangle_{IC}$ for the set of all possible scenarios with only base relations from the corresponding restriction R_{IC}. $[\![\cdot]\!]_{IC}$ stands for the subset of $\langle\!\langle\cdot\rangle\!\rangle_{IC}$ of scenarios that are also consistent. Then we consider the set of all scenarios from $\langle\!\langle \cdot \rangle\!\rangle_{IC}$ which are at least as close to $\mathcal{N}$ as the closest scenario from $[\![\cdot]\!]_{IC}$:

$$S(\mathcal{N}) = \{ s \in \langle\!\langle \cdot \rangle\!\rangle_{IC} \mid \forall s' \in [\![\cdot]\!]_{IC} : d(s', \mathcal{N}) \geqslant d(s, \mathcal{N}) \} \tag{1}$$

The final merging result is the union of all these scenarios:

$$\Delta(\mathcal{N}) = \bigcup_{s \in S(\mathcal{N})} s \tag{2}$$

To solve the optimization problem of computing the actual result of our merging operator, we described an algorithm in Wallgrün and Dylla (2010) that generates scenarios in order of increasing distance until a consistent scenario is found. While the algorithms performs reasonably well in practice (in particular when the distance of the closest consistent scenarios is rather small), the worst-case time complexity is still exponential. To accommodate the ICs, we adapted this algorithm by initializing each constraint with those base relations from the corresponding restriction R_{IC} that are closest to the original relation (e.g., the relation between G and H in Figure 3a is initialized as EC). In addition, when relaxing a constraint by considering conceptually neighbored base relations, we now filter out all base relations not contained in R_{IC}.

4. Conclusions

We presented a general framework for integrating spatial data (quantitative or qualitative) from several knowledge bases. To assure that the information is integrated in the most reasonable way given qualitative ICs, the actual merging takes place on the qualitative level. For this, we defined a distance-based merging approach. Its output is used to combine and adapt the quantitative data accordingly. As discussed, parts of our implementation are still tailored for specific calculi or make simplifying assumptions. Developing more general approaches, in particular ones that are able to deal with multiple QSCs for different aspects of space, will be the goal for future research.

Acknowledgements

Financial support by the Deutsche Forschungsgemeinschaft (DFG) under grants IRTG GRK 1498 SIGI and SFB/TR 8 Spatial Cognition is gratefully acknowledged.

References

Cohn AG and Renz J, 2007, Qualitative spatial reasoning. In: van Harmelen F, Lifschitz V and Porter B (eds), *Handbook of Knowledge Representation*, Elsevier.

Condotta JF, Kaci S and Schwind N, 2008, A Framework for Merging Qualitative Constraints Networks. In: Wilson D and Lane HC: *Proceedings of the Twenty-First International Florida Artificial Intelligence Research Society Conference,* 586-591.

Dylla F and Wallgrün JO, 2007, Qualitative spatial reasoning with conceptual neighborhoods for agent control. *Journal of Intelligent and Robotic Systems*, 48(1):55–78.

Egenhofer MJ, 1989, A formal definition of binary topological relationships. In: Litwin W and Schek HJ (eds), *3rd Int. Conference on Foundations of Data Organization and Algorithms*, 457–472.

Freksa C, 1991, Conceptual neighborhood and its role in temporal and spatial reasoning. In: Singh M and Travé-Massuyès L (eds), *Decision Support Systems and Qualitative Reasoning*, 81–187.

Randell DA, Cui Z and Cohn A, 1992, A spatial logic based on regions and connection. In: *Principles of Knowledge Representation and Reasoning: Proceedings of the 3rd Int. Conference*, 165–176.

Wallgrün JO, Frommberger L, Wolter D, Dylla F and Freksa C, 2007, Qualitative spatial representation and reasoning in the SparQ-toolbox. In: Barkowsky T, Knauff M, Ligozat G, and Montello D (eds), *Spatial Cognition V,* 39–58.

Wallgrün JO and Dylla F, 2010, A relation-based merging operator for qualitative spatial data integration and conflict resolution. Technical Report 022-06/2010, SFB/TR 8 Spatial Cognition.