CLASSIFICATION RULES GENERATION ON FUZZY SPATIAL DATA CUBES

RAJIV KUMAR, Dr S.Venkatanarayanan
[1]Research Scholar, SSSUTMS, Sehore
[2]Research Guide, SSSUTMS, Sehore

Abstract: Addressing the indeterminacy of spatial information from multiple sources is a major challenge in any decision-making process. This applies in particular to the assessment of the risks caused by natural events when decision-making models are based on a range of natural, economic and social indicators. In this context, many of these indicators are semantically vague, others are spatially or temporally uncertain. Most indicators can be assessed at different scales according to the needs of different organizations. Current decision support systems generally do not consider the inaccuracy of information, but assume that indicators have well-defined and accurate semantics, geometry, and temporality.

Keywords: Spatial information, Fuzzy aggregation, Fuzzy Intersection

Introduction

The blurring of information is an essential feature of the data that must be considered in every decision-making process (Kentel and Aral 2007, Bejaoui 2009). Information inequalities can be propagated during the integration and aggregation processes and thus also in the decision-making process. Ignoring inaccurate information can lead to unrealistic or misleading conclusions and decisions that have unintended or even catastrophic consequences. For example, decisions based on unclear environmental, economic and social indicators can lead to incomplete territorial coverage of flood insurance, poorly positioned erosion protection infrastructure and changes to contingency plans, more victims, etc.The information vagueness in assessing the risks caused by natural phenomena can be characterized as semantic, spatial, and temporal. Most of these phenomena have also multi-scale characteristics (Cheng et al. 2009). Several works have already been initiated to deal with information vagueness (Pauly& Schneider 2010; Bejaoui 2009; Edoh-alove et al. 2013; Schneider et al. 2011; Schneider 2010; Schneider 2003a; Fisher 2008). From the technological point of view, existing tools do not provide built-in capabilities to deal with information vagueness either. For instance, in the case of risk assessment, multiple decision indicators are required to define and then translate into datacube dimensions to calculate potential risk.

Fuzzy Spatial Datacube

Spatial datacubes are designed to deal with the cross-tab of hierarchical semantic systems (Bédard et al.2009). The main elements of a spatial datacube (i.e. level's attributes, spatial levels, spatial members, spatial dimension, spatial hierarchy, spatial measures, and spatial facts) are formally defined in Salehi (2009). Data in a spatial datacube can also be uncertain or vague. Moreover, the information vagueness may also arise in the definition of dimensions, hierarchies, aggregation relationships, aggregation functions and spatial measures (Sboui 2010). Thus, information vagueness should appropriately be handled in a spatial datacube. Embedding information vagueness in the multidimensional model requires redefining the principal elements of the spatial datacube which is explained hereafter. In this regard, a fuzzy approach based on Fuzzy Set Theory is proposed and is applied to a spatial multidimensional model proposed for CERA in Jadidi*et al.* (2013).Characterization (i.e., generalization) can be used to generalize task relevant data into generalized data cube. Characteristic rules, which are extracted from a generalized data cube, summarize general characteristic of user-specified data. Similarly, characteristic rules, which are extracted from a fuzzy spatial data cube, can summarize the climate data for a region with the extension that they can also present the general characteristics of the region's climate with some precision. The raw data for one region can be generalized into concepts like cold (0.9), mild (0.7) and hot (0.5) for temperature, and dry (0.28), wet (0.72) for precipitation with the precision values that indicates the degree of reliability of the generalization.

Fuzzy aggregation in spatial Datacube

Aggregation in Geospatial Business Intelligence (GeoBI) community is the grouping of data geometrically, thematically, or semantically to a coarser level of detail (Pedersen et al. 2001; Gomez et al. 2009). This concept of aggregation is very different from the map generalization process (Bédard et al. 2007) and is not the same as in object-oriented modeling (Laurent 2010). In fact, aggregation in GeoBI is a summarization process of values or geometries in a datacube that directly depends on the data model used (Péres et al. 2007; Laurent 2010; Gomez et al. 2009; Pedersen et al. 2001). It typically uses SUM, AVG, MIN, MAX, COUNT and similar operators, but also more complex ones such as spatial operators (e.g. overlay, intersect, include, and fusion), and advanced statistical formula or simulation algorithms. In addition, the geometry used at the different levels of abstraction often comes from different datasets.Using fuzzy concepts to define appropriate operators for data aggregation in a datacube has been initiated by Laurent 2010 and Molina et al. (2006). A series of operators such as roll-up, drill-down, slice, dice, and pivot have been defined for fuzzy datacubes in Molina et al. 2006 and Martin-bautista et al. (2013) using both quantitative and qualitative data. This permits a qualitative representation of results on charts and tables. The thematic aggregation can principally be performed based on Laurent 2010 and Molina et al. (2006). However, the geometric aggregation involving spatially fuzzy or crisp members requires redefining fuzzy operators such as overlay and fusion for fuzzy spatial objects. Spatial relations in fuzzy aggregation follows the ISO standard model (Dilo 2006) and the true/false values of the ISO relations have been extended to a fuzzy degree between 0 and 1 that represents its truthiness. The fuzzy operators are equivalent to their crisp counterparts when applied to crisp objects, since crisp objects have maximum degree of membership of 1.Instead of generalizing spatial data to temperature regions", "[2-3] temperature regions" and then aggregating them to "hot regions"; generalizing each spatial datum to "hot region" with the precision value for the reliability to that generalization μhotand then aggregating them to "hot regions" with a μhot for the aggregated regions is more meaningful and natural. Different temperature values will cause to generalizations with different precision values. Introducing fuzzy logic to spatial generalizations helps to have more smooth generalizations.

Fuzzy Union

Fuzzy union operator results a new fuzzy object from combining multiple fuzzy objects, fuzzy partitions in this paper, with the membership values of whom with higher membership degree (Dilo et al. 2007; Schneider2003a).

Union

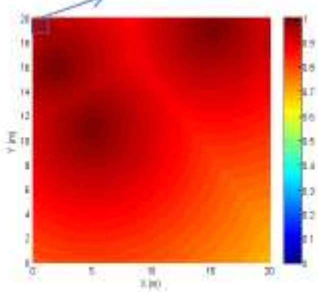
$$FuzzyOverlay : FuzzyPartition \times FuzzyPartition \rightarrow FuzzyPartition$$

$$\forall P_1 = \{\mu_i\}_{i=1}^{n}, P_2 = \{\nu_j\}_{j=1}^{m} \in FuzzyPartition,$$

$$FuzzyOverlay(P_1, P_2) = \left\{ \begin{array}{l} \zeta_{i,j} \mid i \in \{1...n\}, j \in \{1...m\}, \\ \zeta_{i,j} = Union(\mu_i, \nu_j) = Max(\mu_i, \nu_j) = \mu \cup \nu \end{array} \right\}$$

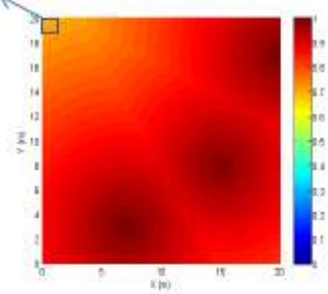
Fuzzy Region 1

1.0000	0.7856	0.7858	0.7861
0.7859	0.7862	0.7864	0.7867
0.7865	0.7867	0.7870	0.7872
0.7870	0.7873	0.7876	0.7878

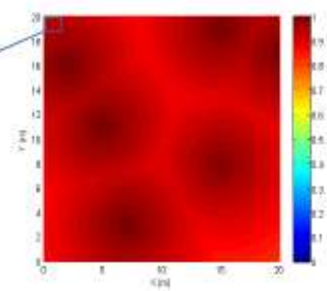


Fuzzy Region 2

1.0000	0.8593	0.8600	0.8606
0.8590	0.8596	0.8603	0.8609
0.8593	0.8599	0.8605	0.8612
0.8595	0.8602	0.8608	0.8614



1.0000	0.8593	0.8600	0.8606
0.8590	0.8596	0.8603	0.8609
0.8593	0.8599	0.8605	0.8612
0.8595	0.8602	0.8608	0.8614



The formal definitions of fuzzy union, example of numerical values and their geometrical representations

Fuzzy Intersection

Fuzzy intersection operator results a new fuzzy object from combining multiple fuzzy objects, here fuzzy partitions, with the membership values of whom with lower membership degree (Dilo et al. 2007; Schneider 2003a). That means fuzzy intersection produces a new fuzzy region by taking the cells with lower membership value inside any overlapping area.

Association discovers a set of association rules in the form of X1 ^ … ^ Xn -> Y1 ^…Ym, at multiple levels of abstraction from the relevant set(s) of data in a database. Association rule discovery necessitates the computation of support to find the frequent item sets and the computation of confidence to find the interesting rule. Computation of these factors requires the count of occurrences of the corresponding item set and the count of all item sets. The data cubes facilitate efficient mining ofassociation rules since a count cell stores the number of occurrences of the corresponding multi-dimensional data values and a dimension count cell stores thesum of counts of the whole dimension. These count cells simplifies the calculation of support and confidence measures of the association rules. But, in fuzzy data cubes,the interesting association rules can be determined according to their significance andcertainty factors, these reflects the reliability to generalization, instead of support andconfidence factors which reflect the frequency of the data.

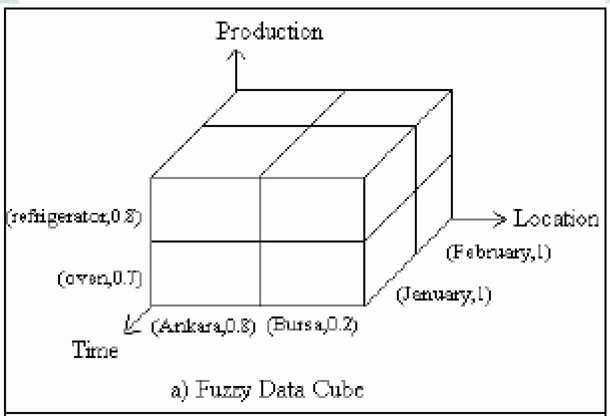
Spatial Generalization

Region	Temperature
R1	Hot
R2	Hot
R3	Hot
R1, R2, R3	Hot

Fuzzy Spatial Generalization

Region	Temperature	Temperature Membership
R1	Hot	0.98
R2	Hot	0.67
R3	Hot	0.56
R1, R2, R3	Hot	?

In Laurent's study [10] for fuzzy cubes, each slice corresponds to the cube with a membership value, i.e., a value of one dimension has the same membership value for all the cells in the slice. But in our fuzzy spatial data cube each cell has its individual membership value for the corresponding dimension value since spatial objects might have common properties but eachspatial object might have that property with a different degree than other spatial objects as displayed in Figure 2.



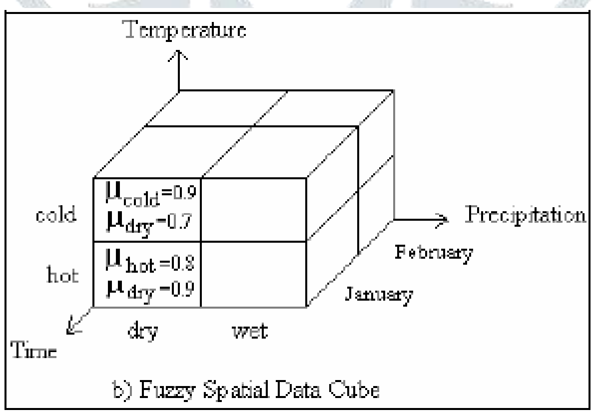


Figure 2. Dimensions and Their Memberships

Conclusion

Fuzzy spatial datacubes are essential to perform more comprehensible knowledge discovery for effective decision-making. An example in this regard is assessing the risks caused by natural phenomena like erosion in coastal regions. A fuzzy-logic-based approach was proposed in this paper to deal with information vagueness originated from the uncertainty of an object and its geometry definition. This concept was then embedded into a spatial datacube through redefining the spatial datacube elements (dimensions, members, hierarchies,measure and facts) as fuzzy dimensions, fuzzy members, fuzzy hierarchies, fuzzy measures, fuzzy facts and required fuzzy aggregation operators (union, intersection, difference, overlay, and fusion). One of the main advantages of using a fuzzy spatial datacube is the capability to present and report the results to end-users using linguistic expressions. Another advantage is representing separately the level of uncertainty and vagueness of the calculated measures for a more realistic decision-making.

References

[1] Abelló, A., Saltor, F. & Samos, J., 2006. YAM2: A Multidimensional Conceptual Model Extending UML. *Information Systems*, 31(6), pp.541–567.

[2] Abuodha, P.A. &Woodroffe, C.D., 2006. *International Assessments of the Vulnerability of the Coastal Zone to Climate Change, Including an Australian Perspective* D. of the E. and H. Australian Greenhouse Office,ed., Australian Greenhouse Office.

[3] Adger, W., 2006. Vulnerability. *Global Environmental Change*, 16(3), pp.268–281.

[4] Aerts, J., Goodchild, M.F. &Heuvelink, G., 2003. Accounting for Spatial Uncertainty in Optimization with Spatial Decision Support Systems. *Transactions in GIS*, 7(2), pp.211–230.

[5] Alexander, D., 2000. *Confronting Catastrophe: New Perspectives on Natural Disasters*, Terra Publishing.

[6] Ali, A. &Rakus-Andersson, E., 2009. Fuzzy Decision Making in Business Intelligence Application of fuzzy models in retrieval of optimal decision. *School of Engineering, Mathematics Dept*, Msc, p.31.

[7] Altman, D., 1994. Fuzzy set theoretic approaches for handling imprecision in spatial analysis. *International Journal of Geographical Information Science*, 8(3), pp.271–289.

[8] Bakillah, M., 2012. *Real Time Semantic Interoperability in Adhoc Networks of Geospatial Databases : Disaster managment*. Université Laval, Qubec City, Canada.

[9] Bédard, Y. et al., 1997. *Étude de l'étatactuelet des besoins de R&D relativement aux architectures et technologies des data warehouses appliquées aux donnéesspatialesResearch*, Quebec City: Lavaluniversity.

[10] Bédard, Y., 1988. Uncertainties in Land Infromarion Systems Databases. In N. Chrisman, ed. *Auto-Carto*. Baltimore, United States, pp. 175–184.

[11] Bédard, Y., Merrett, T. & Han, J., 2009. Fundamentals of spatial data warehousing for geographic knowledge discovery. In H. J. Miller J.H., ed. *Geographic Data Mining and Knowledge Discovery*. Taylor & Francis, pp. 53–73.

[12] ES Kumar, K Vengatesan, RP Singh, C Rajan**,"** Biclustering of gene expression data using biclustering iterative signature algorithm and biclustering coherent column",vol 26, issue 3-4, pp. 341-352,2018.

[13] Vengatesan K., Mahajan S.B., Sanjeevikumar P., Mangrule R., Kala V., Pragadeeswaran (2018) Performance Analysis of Gene Expression Data Using Mann–Whitney U Test. In: Konkani A., Bera R., Paul S. (eds) Advances in Systems, Control and Automation. Lecture Notes in Electrical Engineering, vol 442. Springer, Singapore.