Spatial Databases - Creative Future Concepts and Use

George L. Benwell

Director

Spatial Information Research Centre

University of Otago

New Zealand

Ph + 64 3 479 8300

Fx + 64 3 479 8311

Em gbenwell@commerce.otago.ac.nz

Abstract

There is continuing pressure to develop spatial information systems. This paper develops two concepts that could emerge. The first is a new spatial paradigm - an holistic model - which is less of an abstraction from reality than current models. Second, is the concept of federated databases for the improved and transparent access to data by disparate users. The latter concept is hardly new and is included in this paper to emphasize its growing importance. These two developments are presented after a introductory discussion of the present state of the discipline of geographical information systems and spatial analysis.

Introduction

The spatial databases of tomorrow will be shaped by two major influences - an improved ability to model reality and users demanding increased database access. Other potential influences such as price, platform, speed, storage capacity are considered here to be less important. This paper first outlines the limitations of current spatial databases and database design and then develops a conceptual framework for future design and access.

In the last decade spatial databases have continued to be designed and developed at a frenetic pace. This pace, in relative terms, is slowing - the quantum leap has yet to occur. Current spatial databases have a short history - not much more than 15 years. For the majority of that time they have evolved out of the proprietary pursuits of the leading vendors. Initially, Intergraph and ESRI (and others of less endurance) dominated the field. Both companies responded to market forces and developed systems with simple database structures. The databases may have been flat files but mostly of the network and hierarchical structures. Advances have moved to relational and commercial databases. This has facilitated development and increased accessibility to data.

Spatial data remains separated from attribute data and there remains two distinct structures for representing spatial objects - vector or raster. Topological associations are now incorporated into the data structures but a true object paradigm has not become entrenched. This apparent reluctance to embrace the object oriented structures is intriguing given the early appearance for object oriented GISs (e.g. System 9). The acceptance of full topological and object structures were initially constrained by hardware and secondary storage. This is no longer the case but still there is no common use of object oriented database management systems (OODMSs).

There are considerable demand for improved spatial systems - where will they develop? This can partly be answered by understanding that current systems emerged from geography, geomatics and mapping disciplines. These disciplines, it is contended, are responsible, rightly or wrongly, for the current spatial modelling paradigms. Will pressures from users and business imperatives direct development into areas such as, visualization, spatial analysis, data structures and databases or general functionality?

This paper proposes that development will occur where commerce responds to pressure but that theoretical research should occur in the areas of spatial modelling concepts. It will be important to discard the historical mapping shackles (with due acknowledgment to the valuable but constraining contribution). A (r)evolutionary spatial modelling paradigm awaits discovery.

An improved spatial model

Where will it come from - what will its concepts be founded on - how will it be revolutionary? If it is possible to visually perceive the world as it surrounds us, then it logically follows that it is possible to mentally store and recall these images. The human mind is capable of observing and understanding and even analyzing space and spatial objects.

Spatial databases will evolve to store and retrieve representations of space where these representations are more accurately (closer to the truth) aligned with the phenomena we perceive. It is tempting to delve into the workings of the mind, though this is resisted here and now. What is more important to realize is that current databases are constrained by their own pedigrees. In future, space and objects within it, collectively need to be represented differently in databases. Saying 'space and objects' is already pre-emptive. It is a break down of convenience; it may not be the best form. It is used here simply as a convenient way to explain an emerging concept.

It is held and developed here that spatial databases of the future will store space and objects as a whole - just as we perceive the world around us. Raster and vector will give way to space, object and knowledge. Representations will be born out of the concept that objects exist and have knowledge embedded in them - that they exist and behave in an observable and representative way. This may seem to be much the object oriented paradigm. It is more than that - much more if we are to succeed.

Figure 1 indicates the de-aggregation of reality and its re-aggregation into representations that may be stored in a database. The last two lines are a casual (or somewhat flippant) comment on possible scenarios. Any developments in spatial modelling must understand the primitive and substantive reason why user want to model that environment - space.

Object = Reality - Abstraction

Data = Object + Measurement

Information = Object + Data + Structure

Knowledge = Information + Use

Wisdom = Knowledge + Experience

(Hopelessness = Data + Experience)

(Hopefulness = Objects + Knowledge)

Figure 1

It is intuitive that users collect and store spatial to either present it as is or to analyse it. The underlying reasons are possibly very complex and what's more possibly not all that relevant to this discussion. It more important to recognize that users are interested in understanding how components work together or the processes that have created and continue to modify a large continuous spatial phenomenon. A good example of the former is how people and albatross interact at a tourist attraction (reported elsewhere at this conference and Purvis et al., 1993) or why of two forest trees, why one is taller and healthier than the other. In the second case, is it possible to explain a geological fault, why a city exists, or what drives ocean upwelling and saline fronts.

From these isolated but typical examples it is not unrealistic to draw a conclusion that what is required is a model of reality, as intricate as possible, that relates a suitable number of its de-aggregated components, and allows for the analysis of these components. There is no use creating a wonderful model if the components are not identifiable or their interactions obliterated.

With this an explicit background it is possible to develop Figure 2. This figure shows the abstraction of reality data, entities and space. The implication is that initially the abstraction is to data as a consequence of the current field collection methods and techniques. When data are appropriately re-aggregated, eventually *Space* and *Rules of existence* are (re-)created. The choice or use of words here is somewhat arbitrary but nonetheless important. *Object* is not used so as to avoid confusion, but *entity* is unavoidable. The words used in Figure 2, may to some, abuse convention; the intention is that the higher one ascends in the text, the more holistic the model becomes.

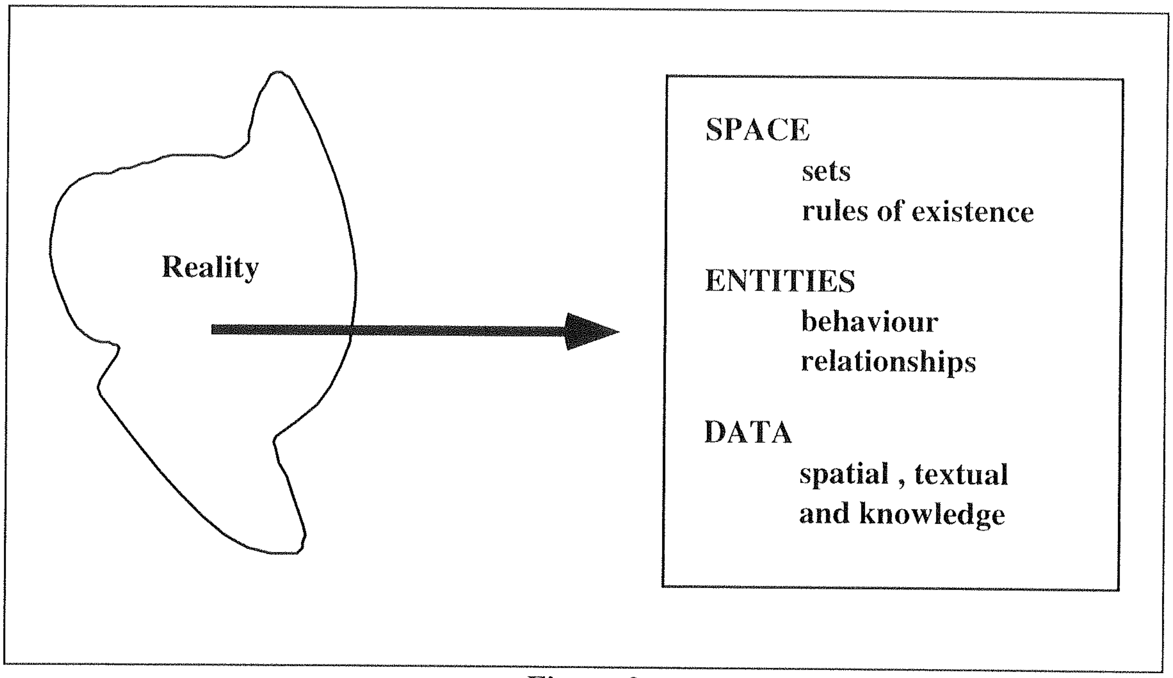

Figure 2

Current systems are at the entity (again, not to be confused with database terminology), behaviour and relationships level. This may be manifest in relational databases, or object databases, either with full topological data structures.

The *Space* stage with *sets* and *rules* may too be represented in a relational database. While that construct may be the same, the users' fundamental view of the data are different. What is seen is a set of spatial phenomena and a minimum set of rules that describe their classification. This could run in parallel with an object paradigm which stores these phenomena and their associated inheritance, behaviours etc.

If it is possible to store this structure then retrieval will be beneficial; Frawley *et al.* (1992) described similar situations as the discovery of knowledge or knowledge extraction. They went on to say that what was really required from databases was, *nontrivial extraction of implicit, previously unknown, and potentially useful information from data.* This in turn required a special classification technique for the phenomena being studied and means of retrieving or *mining* through the database.

To return to the human mind and perceptions of space for just a moment. The system description given above could be developed to be similar to the way we understand human discovery algorithms. We have a range of different methods for different types of problems. In fact we have different strategies for learning different problems, especially for classification, which apply a number of different methods to the same task and select rules from the best method (Brodley, 1993; Michalski and Tecuci, 1993).

These concepts can be considered to be closely related to the fields of rough sets and database knowledge discovery. The theory for the former was originally developed by Pawlak in the early 1980s. The primary methodological framework was to study classification problems with imprecise or incomplete information and the theory of rough sets was developed. This theory, to the best of the author's knowledge, has not been implemented and tested with spatial phenomena and knowledge extraction. There is considerable appeal to apply the theory to spatial data.

If the high level model shown in Figure 2 can be developed it may well be achieved using rough sets. The concept would be to develop a database of spatial phenomena, where the unique (or nearly unique) classification of the components is achieved with a minimum group of rules or class definitions. The key idea in the rough sets approach stems from the observation that imprecise representation data helps uncover data regularities. Any concern or belief here that we in fact know precisely the data and relationship should be discarded. The point of most geographic information systems is the analysis of the data to determine exactly that - what are the data regularities (and hence may be data irregularities)?

The theory of rough sets provides a collection of mathematical techniques to deal, with full mathematical rigour, with data classification problems (which once we would have called a layer or theme), particularly when data are noisy, incomplete or imprecise. Also rough set theory includes a formal model which defines knowledge as a family of indiscernibility relations which gives knowledge a clearly defined mathematical sense. This definition allows 'knowledge' to be analysed and manipulated using mathematical techniques (Ziarko, 1993).

Again to return to a previous discussion. What is the purpose of a geographic information system and how are the data derived? The former has already been discussed. For the latter, the usual method is to employ a method which collects [x,y,z, attribute] data on a pseudo-random (feature driven tessellation) or a uniform basis (raster regular tessellation). Knowledge collection is either held in the mind of the data collector or is disregarded and forgotten. The result is an incomplete or imprecise situation from which the spatial scientist desires to determine spatial relationships and knowledge.

For those wishing to examine the theory of rough sets in more detail refer to (Pawlak, 1991). The brief material presented here is only to demonstrate that an high level spatial modelling paradigm, as shown in Figure 2, is both conceptual possible and logically and spatially sound.

Federated databases

Hand in glove with database developments is a concept of federated databases. This is, in simple terms, the connection of several databases in such a way that the union is transparent. Furthermore applications can access an data repository index which contains information on data, data structures, data use rules, data knowledge and other applications. The federation makes this possible from any application domain to any other using any data for any reason. A grand concept that is close to reality.

The concept of federated databases seems to be remote from the concept of rough sets. That may well not be the case. For, if there is a desire to extract knowledge from disparate data sets using the techniques described above, one of the imperatives will be easy access. It must be possible to trawl through several disparate (both in the contextual and geographic senses) databases in a way that does not inhibit the user of the search algorithms.

As an example consider a client - a private electricity company which relies on hydroelectricity to meet peak load demands. Such a company will have to manage its resource, water at altitude and have appropriate models of supply and demand. These will be very variable in both time and location. Applications would include, socio-demographics, climate and rainfall models, snow melts and load determination. It takes little further imagination to list the data sets required for these applications and that there will not be one single data source.

Future federated databases will provide an environment that execution of the models will be possible without detailed knowledge of where the data are or what structure they are stored in. It will even possible to remotely bill your account - of course.

Conclusion

This paper set out to introduce an alternative paradigm for spatial modelling. The theory of rough sets is, at least prima facae, seen to have the ability to address this concept, although there is no known application for spatial data at this time. Work by Colin Aldridge at Otago is showing encouraging developments. The paper concludes with a short discussion on federated databases and the access to disparate databases. Both these topics, rough sets and transparent access to data via open systems, are looming large on the spatial research horizon.

There is also considerable research to determine the interactions with rough sets and geographical analysis techniques such as case based reasoning, statistics, kriging and the like.

Acknowledgments

I wish to thank the researchers in the Spatial Information Research Centre for their help in getting us all to Leeds. I thank them for the assistance in preparing this paper. On a particular research note, I acknowledge the in-depth discussion with one of my PhD students, Colin Aldridge, whose persistent and inquiring mind uncovered the connection between rough sets and spatial phenomena.

References

BRODLEY, C., 1993. Addressing the Selective Superiority Problem: Automatic Algorithm/Model Class Selection, In Proc., of 10th Machine Learning Conference 17-24, Morgan Kaufmann.

FRAWLEY, W., PIATETSKY-SHAPIRO, G. and MATHEUS, C., 1992. Knowledge Discovery in Databases: An Overview. *AI Magazine*, Fall issue.

MICHALSKI, R. and TECUCI, G., 1993. Editors, *Machine Learning: A Multistrategy Approach*, Volume IV. Morgan and Kaufmann.

PURVIS, M.K., GASKIN, C., SMITH, I., and McLENNAN, B.R., 1993. Life at Taiaroa Head (Pukekura). *Fifth Annual Colloquium of the Spatial Information Research Centre*, Proceedings. Eds. BENWELL, G.L. and SUTHERLAND, N.C. Dunedin, University of Otago, New Zealand, pp265-278.

ZIARKO, W., 1993. Rough Sets and Knowledge Discovery: An Overview, In *Rough Sets, Fuzzy Sets and Knowledge Discovery*, Workshops in Computing, Ed. C.J. van Rijsbergen, pp11-15, Springer-Verlag.