roar @UEL
research open access repository

University of East London Institutional Repository: http://roar.uel.ac.uk

This paper is made available online in accordance with publisher policies. Please scroll down to view the document itself. Please refer to the repository record for this item and our policy information available from the repository home page for further information.

To see the final version of this paper please visit the publisher's website. Access to the published version may require a subscription.

An Agent-Based System to support Geo-Information Analysis

Abdullah Al-Zakwani, Haralambos Mouratidis and Allan Brimicombe

University of East London
{A.Al-Zakwani, H.Mouratidis, A.J.Brimicombe }@uel.ac.uk

Abstract - Geo-Information Systems (GIS) are software for the handling and analysis of spatial data and are at the heart of Geo-Information Science (GISc). Most GIS functions require interaction with human experts. This is considered problematic since it causes extra complexity and increases the amount of resources required. Agent Technology has the potential to assist in reducing this problem. However the current application of software agents to the GIS domain is very limited and fails to take into account the full functionality and advantages of Agent Technology. In this paper we discuss the application of agents in GIS and argue for the need to produce an agent-based framework for GIS. By way of context we define a novel agent-based system for one important aspect of GIS, the construction of a variogram. We quantitatively compare our architecture with two the existing agent-based tools: RePast for ArcGIS and Oracle Agents.

1. Introduction

Geo-Information Science (GISc) is the field of study that addresses substantive questions concerning the handling, analysis and visualisation of spatial data. Geo-Information Systems (GIS) are a key technology for GISc though they are often used in conjunction with other technologies to form geo-computational environments or core support for location-based services (LBS). GIS-related scientific questions concern issues of data quality, geostatistics (defining spatial dependencies), visualisation, geosimulation for a range of applications. Given the breadth of natural, social and economic phenomena to which GIS can be applied, the expertise of the user can play a pivotal role in successful analyses [1]. This is especially true for Geostatistics, a branch of GISc concerned with the analysis of spatial data and most specifically used in connection with defining and measuring spatial dependencies in data and applying this in spatial interpolation (contouring). A key tool in measuring spatial dependencies is the variogram which has the general form [2]:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N} \left(z(x_i) - z(x_j) \right)^2 \qquad (1)$$

...where h is spatial lag, N is number of data points, z is the variable being modelled as determined at points x_i and x_j.

The process of constructing the variogram (Fig. 1) and modelling the spatial dependencies usually requires heavy intervention of an expert (human) user. For example, there may be a need to assemble data from across a network, a requirement to understand the nature of the data and its context, and to fit an appropriate model (e.g. Gaussian, spherical, quadratic; plus any anistrophic effects). Tool interoperability is also an important consideration in variogram construction and other aspects of geostatistics.
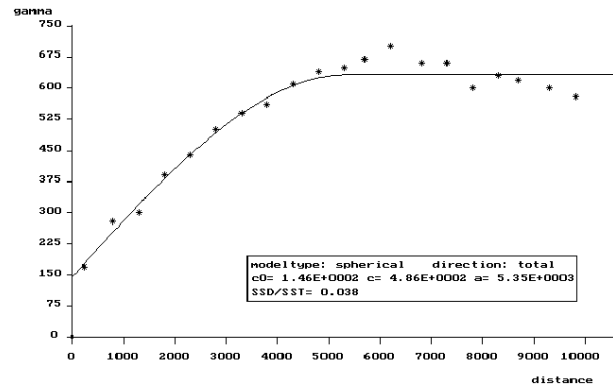

Fig. 1. An example of a fitted variogram model.

Artificial Intelligence (AI) has been extensively used to analyse geographical phenomena [3]. In particular, many features of AI like neural networks and fuzzy sets have already been used in conjunction with GIS [3] [4] [5]. Such approaches, however, are not particularly suited to heterogeneous data environments with low levels of inter-operability particularly across networks. This is important since GIS tools are made increasingly available through the Internet, and are being ported to mobile networks for LBS applications [6]. Consequently the issue of interoperability urgent needs to be addressed [7] [8].

Recently the spotlight has fallen on agent technology as a means of solving the above problems. However, from our analysis, we feel that the current application of agents in GIS is not consistent with the understanding of the concept of software agent in the agent community [9]. In particular, within the GIS community, the concept of an agent is used as a static entity to study emergent behaviour *in silico* as an aid to simulation [10] [11]. However such treatment of software agent fails to realise the full potential of the technology. Our research is focused on investigating novel

ways of using agents in GIS where a primary requirement of the applications is the acquisition, verification and dissemination of knowledge between agents. In this paper we outline an agent-based system that deals with the construction and modelling of the variogram, as a significant challenge within GIS. We go on to present a quantitative comparison of one aspect of the system against two existing GIS tools that claim to incorporate characteristics of agent technology: RePast for ArcGIS and in Oracle Spatial.

2. The variogram agent-based system

The proposed system has been analysed and designed with the aid of the Tropos agent oriented software engineering methodology. The agents themselves have been constructed using JACK (www.agent-software.com). The main actor is the geostatistician requiring construction and modelling of a variogram. Decomposing the system to a set of agents each responsible for satisfying the system goals and tasks results in the following agents (Fig. 2):

- **Data Finder**: responsible to acquire data, analysing its structure and providing it to the rest of the system.
- **Integrity Checker**: responsible for determining errors and cleaning the data.
- **Data Analyser**: responsible for determining if the type of data has been received before or is a new data type. Where appropriate it will de-trend the data.
- **Sampler**: responsible for determining sample sizes for modelling and integrity checking of the variogram.
- **Mathematical Modeller**: responsible for applying the appropriate formula for the calculation of γ given a particular data structure.
- **Strategy Comparer**: responsible for checking the results produced by Sample and Mathematical modeller and determine the accuracy.
- **Model Fitter**: responsible for checking the plots and for providing a best-fit curve as a model.

Our investigation in this paper focuses on the Data Finder agent. Fig. 3 models the goals/tasks of this agent. As mentioned above, the Data Finder agent is responsible for acquiring data and providing data to other agents in the system after finding their structure and type. The source of the data could be a human input, remote sensing portals, or other computing or distributed devices. Moreover, the Data Finder agent depends on the Integrity Checker agent and Geostatistician (the expert user) in order to satisfy some of its goals.

3. A quantitative comparison of the system

Some aspects of agent technology have already been used with GIS [7] [8]. Two key developments have resulted in GIS tools: RePast for ArcGIS and in Oracle Spatial.
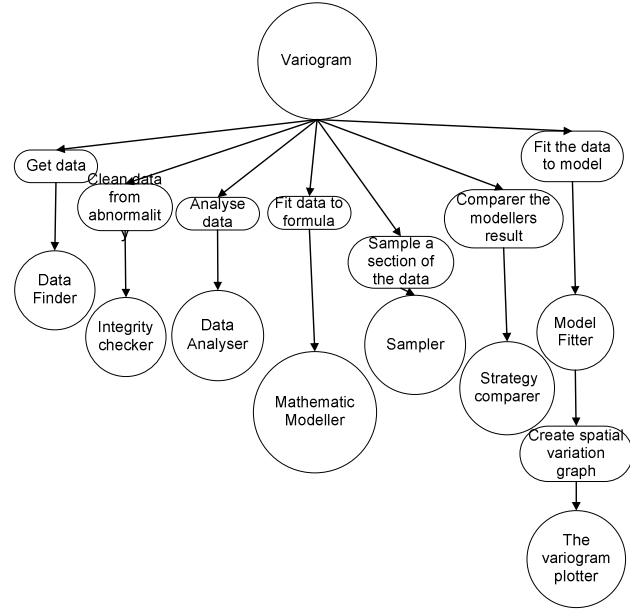

Fig. 2. Decomposition of the variogram system

Fig. 3. Goal analysis for DataFinder agent.

In Oracle, agents are represented as java components (classes and interfaces) with specific tasks. The terminology employed by these "agents" is consistent with the software agent community; however, their functionality and implementation provide a structure such that they are autonomous and reactive but nevertheless would be defined as weak agents [9]. Naude and Chandrasekar [12] define the Oracle

Intelligent Agent (OIA) as "an autonomous process that needs to run on a remote node in the network to make the node Oracle Enterprise Management (OEM) manageable". However their definition determines static objects that help connecting a client's request using a unified gridded server. Though these static entities exhibit one of the very important and powerful characteristics of agents – autonomicity - they lack important functions like social ability and intelligence that could play important roles to determine agents as the role players of GIS analysis and functionality. Nevertheless, Oracle is one of the very few tools that relate agent technology as dynamic entities that can represent a process.

Recursive Porous Agent Simulation Toolkit (RePast) is an extension of ArcGIS. Gilbert and Bankes [13] define it as "an agent modelling tool that offers simplicity to simulate, develop and execute applications of social nature". In RePast agents are viewed as social actors, permeable, interleaved and mutually defining with cascading and recombinant motives. The agents are implemented using multiple languages including Java, Python, C# and the support of the .NET framework. An agent in RePast has beliefs and desires.

Neither of the tools just described provide the ability to assist in data analysis nor provide ideas on where in the data the extra analysis is required and thus provide very little help to the data analyst expert. The main difference to the system proposed in this paper is that the agents are the actual processes that provide services. By services we mean they help to actually define the processes that should be taken to analyse the data with as little intervention from expert users as possible. A comparison can however be carried out between our Data Finder agent and equivalent functionality by agents in Oracle and RePast.

4. The test environment

The test run made use of the Walter Lake data set that can be found in Isaaks and Srivastava [2] for comparison of analysis. The data has possible clusters and errors that need to be analysed before any GIS can be applied to it if accurate results are to be anticipated. The system we developed comprises a number of collaborating agents where each agent performs autonomously. Thus to be able to accurately compare ours with the other two agent based systems, the test will only use our Data Finder agent as its functionality is very similar to that supported by Oracle and RePast. The initial test will be to measure the time taken to complete to produce the results. The assumption here is that faster is better. We will then examine the accuracy of the results according to those provided by Isaak and Srivastava [2]. This will lead to an examination of the agents' functionalities in each system and the query complexity. The functionality measurement will comprise of the amount of

features the agent possesses (the agency), the data acquisition process and the querying process. The query complexity will look at the amount of knowledge required from the expert to construct the query. This test is not intended to exhaustively examine the result as in GIS different analyses can lead to different outcomes. For this reason we refrain from putting emphasis on the specific outcome but simply compare them to those provided by the expert user.

5. The test and results

In our experiment we were not able to utilise the agents to provide any help in data analysis but we were able to utilise agent for distributing the data. After the query was executed the agent was able to send the results to different node (agents) execution. This allows parallel execution when needed and indeed helps on the speed of the execution run (the analysis). In the variogram agent-based system, when the Data Finder agent is in running mode (executing), the data flow is not only being handled internally it is also communicated to the human expert. The human expert view of this data is shown Fig 4. The data were found to be bivariate and each variable was found to have clusters. The data were divided into univariate sets, with errors and clusters within the data identified.
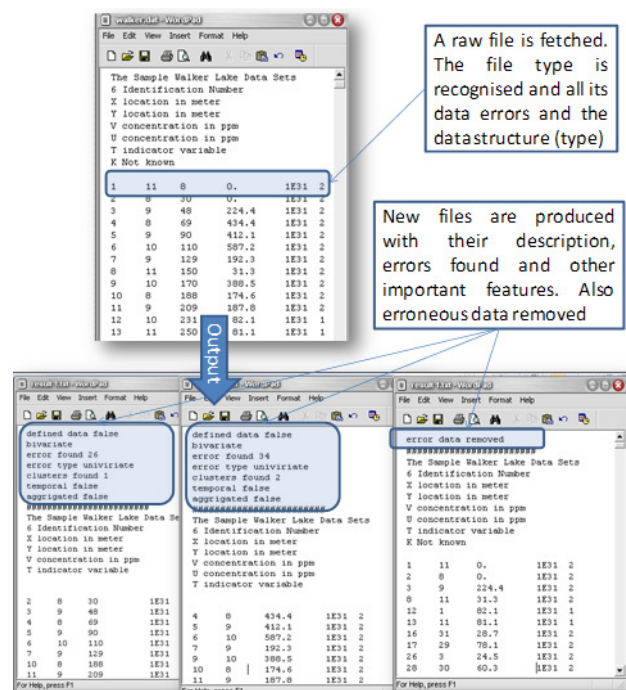

Fig. 4. The exhibition of Data Finder agent.

For Oracle, after the agent mechanism was enabled, a select SQL was used to attempt to minimally analyse the

data. Thus the measurement of the time taken and the results of the query were recorded. The results are shown in Fig. 5(a). The same data analysis was carried out using RePast on ArcGIS. It was only possible to examine the errors by simulating the environment. This is the only feature provided by RePast. However it provided the possibility of seeing the errors. The results are shown in Fig 5(b).
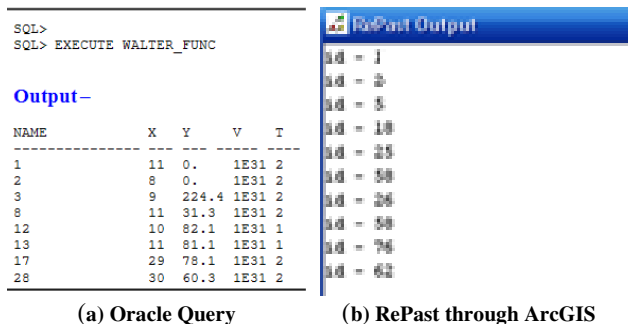

(a) Oracle Query   (b) RePast through ArcGIS

Fig. 5. The results from (a) Oracle and (b) RePast.

The overall results of the investigation are given Table 1. The results show that it is possible to achieve similar results to that produced by our Data Finder agent, however this involves constant interaction with the expert user. It also adds its own complexity as the expert user needs to also be a SQL query designer, as the query we used here is complex PL/SQL with a number of nests. The query process took 3.1 seconds which is a less then a second slower than our Data Finder agent. This difference is negligible except that it takes over an hour to prepare the query for the Oracle while the Data Finder agent requires no query at all. The authors conclude though that the agents in Oracle are useful for distributed data environment particularly in heterogeneous operating systems but not facilitated to deal with intelligence, mobility or any other software agent characteristics.

Table 1. the comparison of agent runs.

Technology	Seconds	Measured in a scale of 10		
	Run to completion	Agent functionality	Accuracy of the results	Complexity of query structure
Oracle	3.1	3	3	10
RePast (ArcGIS)	6.4	3	5	4
Data Finder	2.3	6	7	1

From the perspective of developing agent-based systems for GIS, this test produced satisfactory results particularly given the fact that any one geographical problem can be tackled in a number of different ways. Each expert tends to take their own approach to analysis though hopefully converging on the same outcomes and decisions.

6. Conclusion

In GIS this agent framework and architecture offers flexibility to the user. It offers the capacity of easy extensions, as agents are able to move from one system to another if more resources are required. Also it is learning from the user so that in time the agent becomes expert and requires less and less intervention from the user to solve problems. Given the range of data types and contexts in which the agent system might be applied to, this is the main advantage of applying agents as processes and service providers rather than static objects (entities) as tends to happen currently where GIS use agent technologies.

References

[1] Brimicombe, A.J.: *GIS, Environmental Modelling and Engineering*. Taylor and Frances, London (2003)
[2] Isaaks, E. and Srivastava, R.: *An Introduction to Applied Geostatistics*. Oxford University Press, Oxford (1989)
[3] Openshaw, S. and Openshaw, C.: *Artificial Intelligent in Geography*. Wiley, Chichester (1997).
[4] Brimicombe, A.J.: Encoding expert opinion in Geo-Information System: a fuzzy set solution. *Transaction in Internal Land Management* (2000) 1: 105 – 121
[5] Guesgen, H. and Albrecht, J.: Imprecise reasoning in geographic information systems. *Fuzzy Sets and Systems* (2000) 113: 121 – 131
[6] Brimicombe, A. and Li, Y.: Mobile Space-Time Envelopes for Location-Based Services. *Transactions in GIS* (2006) 10(1): 5-23
[7] Brown, D., Riolo, R., Robinson, D., North, M. and Rand, W.: Spatial process and data models: toward integration of agent-based models and GIS. *Journal of Geographical Systems* (2005) 7: 25 – 47
[8] Reitsma, F. and Albrecht, J.: Implementing a new data model for simulating processes. *International Journal of Geographical Information Systems* (2005) 19(10): 1073-1090
[9] Wooldridge, M. and Jennings, N.: Intelligent agents: Theory and practice. *The Knowledge Engineering Review* (1995) 10(2): 115–152.
[10] Batty, M.: *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals*. MIT Press, Massachusetts (2005)
[11] Albrecht, J.: A new age for geosimulation. *Transactions in GIS* (2005) 9(4): 451-454,
[12] Naude, F. and Chandrasekar, V.: Oracle Enterprise Manager and Management Server FAQ. http://www.orafaq.com/faqoem.htm accessed on 12/023/07
[13] Gilbert, N., and Bankes, S.: Platforms and Methods for Agent-based Modeling, *Proceedings of the National Academy of Sciences of the USA*, National Academy of Sciences of the USA, Washington, DC, (2002) vol. 99 suppl. 3: 7197-7198