An Approach to Online Geoinformation Modeling

Valeri Gitis
Institute for Information Transmission Problems,
Moscow, Russia
gitis@iitp.ru

Boris Osher
Institute for Information Transmission Problems,
Moscow, Russia
osher@iitp.ru

Anton Dovgyallo
Institute for Information Transmission Problems,
Moscow, Russia
dovgyallo@iitp.ru

Tamas Gergely
Applied Logic Laboratory,
Budapest, Hungary
gergely@all.hu

Abstract

The approach to creation of information technology for on-line intelligent geoinformation modeling and complex analysis of space-time properties of geological environment is suggested. The basic client-server architecture of this technology is discussed. The first version of the system that realizes this architecture is described. The main functions of the system are the following: problem formalization, formalization of the information space of the problem, preliminary data processing, generation of a set of features, solution generation, analysis and argumentation of the solution.

1. Introduction

The widespread of GIS set the stage for creating large databases with geographically related space-time information. The development of communication technologies has opened up to Internet users strong possibilities for distributed data access. However the ultimate goal of a user is not only in data search and extraction via Internet as such, but to solve concrete problems [1, 2].

It is necessary therefore to integrate communication tools with network geoinformation environment, that can support in search and extraction of relevant data in the net as well as their cartographic representation, processing, complex analysis and knowledge acquisition.

The development of network geoinformation technologies is most urgent due to with high pace of creating the information society. Such a network environment must be oriented to users demand. Therefore the environment should be portable, so that its functions are extended according to the user's requirements.

Most users solve problems with complete information. Such tasks include general cartographic search, geodata screening and significant part of navigation. Searching, extraction and cartographic analyses of data are usually enough for this category of users. Professional users however, as a rule solve the problems with the incomplete information. These are the problems of investigating and forecasting space-time properties of geological environment such as natural hazard assessment, estimation of ecological processes and zonation, estimation of social and economic risk of natural and non-natural catastrophes, mineral, oil and gas exploration, geoeconomical, geosociological and geopolitical analysis, optimal navigation etc. To solve such complex problems one needs intelligent tools of space-time data processing, analysis, reasoning and argumentation.

The paper outlines GeoNet, the project on WWW on-line geoinformation technology oriented on complex analysis of properties of geological environment and on the decision of geological and geophysical forecasting problems. The technology develops the approach, realized in the geoinformation systems GEO, GEOTIME and GEORISK [3, 4, 5]. GeoNet technology is intended to support the following basic functions:

- data search and extraction,
- cartographic analysis of the relevant information in distributed databases,

Proceedings of the Workshop on Computer Science and Information Technologies CSIT'99
Moscow, Russia, 1999

- processing, complex analysis and argumentation on geographically related space-time data,
- plausible reasoning on space-time objects (hypotheses generation and testing, argumentation of a solution),
- creating the information models of space-time phenomena and processes.

2. Conceptual basis

The main purpose of the GeoNet technology is to support a wide range of users' demands concerning complex analysis and processing of all types of space-time geological and geophysical data with the help of virtual intelligent environment accessible through Intranet and Internet. The GeoNet technology supports two types of users' queries:

(a) Getting knowledge about geological properties of a certain region on the basis of available space-time geodata;

(b) Investigation of properties and solving practical problems of geological and geophysical forecasting.

Tasks related to acquaintance with geological environment are the problems of complete information. The solution process of these tasks is supported with

- the tools of standard statistics on space-time rasters,
- set of the operators for space-time data conversion into the raster fields,
- cognitive cartography, including compiling the maps of similarity with the precedent on the basis of a set of features and compiling the maps on the users' queries, submitted in the form of fuzzy logic constructions.

Investigation and prediction of space-time properties of a geological environment are most complicated both from the point of view of formalization and structuring of the initial information and from the point of view of its further processing and analysis. The complexity is caused by the absence of mathematical models for dependence of numerous geological processes and phenomena on properties of geological environment as well as by incomplete information and noise in initial data.

The high degree of uncertainty dictates the following conceptual scheme of problem solution. On the one hand, there are data and knowledge of a problem domain. On the other, there is an assumption in the form of causal model, about possible relationship between property under consideration with the characteristics of geological environment. The problem is to find a formal solution in the framework of the model. This formal solution must reveal how the studied property depends on characteristics of geological environment and represent its cartographically. Every version of this formal solution is a hypothesis generated as a result of joint processing of available data and knowledge. It is assumed that a solution does not contradict the model if the following three requirements are fulfilled:

- Compatibility: the result is not refused by the available data,
- Significance: the criteria of the quality of the solution brings out the existence of a cause relation,
- Interpretability: the result can be explained and confirmed with the help of the existing knowledge and data.

The foundation of the GeoNet technology contains two methods of reasoning constructions, which are used by specialists in geological and geophysical forecasting. One of them is the construction of logical statements about a relationship between the studied property of a geological medium and the features. The other is the forecast of characteristics of the medium by analogy. The application of these methods is defined by a type of the features used for the description of geological environment.

The characteristics of geological area can be classified as descriptor and indirect features depending on the context of the problem to be solved. A feature is said to be descriptor if its relationship with the property under study is known. E.g. in the case of forecasting oil and gas resources geometrical notions of structural complexes, that are used in the models to estimate the reserve of carbohydrogenes are such features. In the case of seismic hazard assessment the physical-mechanical characteristics of the earth crust are such features.

Characteristics of an area are represented more often by using indirect features. E.g. gravity anomalies are associated with substance and structural heterogeneity of geological area. But this information explicitly does not exist in the known models of carbohydrogenes or in the seismic hazard assessment .

In some cases the indirect features can be transformed into descriptor ones by the use a beforehand given set of operators, submitted in a problem-oriented knowledge base. There is a set of operators in GeoNet for those indirect features the transformation of which is not represented in the knowledge base. These operators ensure the transformation of vector, point and raster data into raster various raster fields features. These fields are structured according to the context of events of interest. For example, fields of density of lines or points, fields, calculated as a result of raster filtration, and also as a result of algebraic and logical operations on several raster fields can be generated.

The solution process of the tasks of analysis of space-time data with incomplete information is iterative. In order to obtain a result a user should as a rule repeat many times the processing and analysis of data, check a set many competitive hypotheses and evaluates the results with respect to the possibility to construct their geological-geophysical interpretations. The versions of the solution are modified and improved according to the accumulation of new observations and getting new data.

The solution process of the investigated problem comprises the following six constituents.

(i) *Formalization of problem:*

- Definition of a way to represent the considered object and/or its properties with the help of some scalar value and a way to describe the object by a set of geological and geophysical features,
- Formulation of a causal model of the phenomenon under consideration,
- Formalization of the model in the form of some cartographic schema and/or in the form of a set of the logical statements.

(ii) *Formalization of the information space of the problem:*

- Selection and structuring of the initial data,
- Representation of geological structures and fields, catalogues of points and time series measured in irregular grid points of a region in vector, raster and table formats,
- Formation of precedent sample sets,
- Selection of operators to transform initial data into raster features that describe the properties of interest in more adequate form,
- Representation of expert knowledge about dependency of investigated property on the features in the form of restrictions on the domain of definition of the parameters to be estimated.

(iii) *Preprocessing:*

- Preliminary processing of linear structure schemes, point catalogues, rasters and time series measured in arbitrary points of a region.

(iv) *Feature generation:*

- Creation of 2D raster fields of linear objects,
- Creation of 2D and 3D raster fields of points objects,
- Linear and nonlinear filtration of 2D and 3D raster fields,
- The execution of user defined algebraic and logical transformations on several raster fields,
- The execution of transformations in view of thesaurus of problem domain to calculate descriptor features.

(v) *Solution generation:*

- Plausible inference of the relationship between the property under study and raster features (forecast function),
- Selection of the most informative raster features,
- Compiling the forecast map and the map of the solution accuracy.

(vi) *Analysis and argumentation of the solution:*

- Error analysis and testing of the interpretability of the solution on the control sample set of precedents,
- Cartographic analysis of the forecast,
- Statistical analysis of the forecast,
- Analysis and confirmation of the forecast by precedent,
- Creation of the logic-based argumentation.

3 GeoNet technology

Internet geoinformation technology GeoNet is based on problem-oriented *information modeling environment*. The components of the environment are the following:

(i) *Information filling of a problem*: contextually formalized and structured data and knowledge,

(ii) *Information contents of the solution*: features, preliminary hypotheses and result,

(iii) *Tools for contextual-dependent formalization and structuring* of data and knowledge,

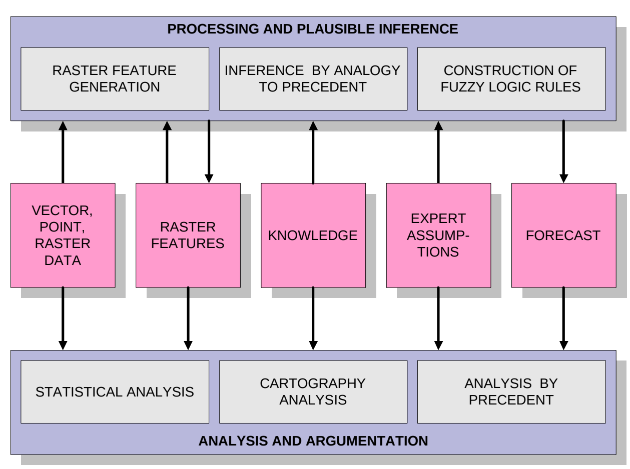
(iv) *Tools for data processing, decision making and knowledge acquisition*,

(v) *Tools for analysis and argumentation* of the derived formalisms.

In an effort to research a problem, to make a decision, to explain and confirm the solution the environment enables to integrate all the information on the investigated phenomenon: multidisciplinary data, operators for initial data transformation in descriptor features, sample sets of precedents, initial knowledge and knowledge acquired during solving a problems, as well as nonverbal expert knowledge.

The interaction between the elements of information environment is submitted at the Fig.1.

Figure 1 Interaction between the GeoNet environment



elements

The tools for data processing and plausible inference allow to realize the expert assumptions about

- ways of formalizing and categorizing the initial data and knowledge,
- techniques to transform initial data in raster features,
- procedures to derive a solution from data and knowledge,
- dependence types of the property under study on raster features.

For *raster feature generation* a wide class of the operators on point catalogues, linear objects, time series, 2D and 3D raster fields is used. With the help of these operators, for example, the following raster features can be calculated:

- weighed density of earthquakes and other seismic regime parameters,
- density of geological faults and lineaments,
- space-time models of geomonitoring data,
- fields received with the help of linear and nonlinear filtering operators,
- fields received with the help user defined algebraic and logic operations on several initial rasters.

Finally all types of initial data are transformed in the raster features which are more adequate to the problem.

Finding a relationship between the property under consideration and features and compiling a digital model of the property are based in technology GeoNet on the construction of logic statements about descriptor features of geological environment and/or on a decision making based on analogy to precedent.

The constructor of fuzzy logic rules is used in GeoNet to realize the decision rules in the form of fuzzy logic expressions. For *inference by analogy* to precedent nonparametric decision making methods of nearest

neighborhood, similarity functions, regression [8] , as well as locally-parametric methods of approximation of interval expert evaluations and likelihood ratio approximation are used [7].

The tools for analysis and argumentation support statistical analysis of the solution, cartographic analysis, analysis by a principle of precedent and construction of logic explanations and confirmations.

The tools of the statistical analysis allow to estimate the basic statistical parameters for the catalogues of point and raster data. They include the correlation analysis, cluster analysis, estimation of expectations, variances and histograms for the raster data of all region and for a set of user defined polygons. Comparing these histograms on the screen allows to explain the properties of selected area in relation to properties of all region under consideration.

Cartography includes tools for data representation and analysis, such as,

- composition of the maps involving several raster, vector and point layers,
- reading the values and attributes of the map elements,
- enlarging a map with or without interpolation,
- the construction of cross sections by user defined profiles for several raster layers with drawing a point depth distribution,
- selecting polygons and sample sets for plausible reasoning, animation of space-time processes, analysis by analogy to precedent.
- compiling the maps of similarity with a selected points by a set of features in the metric l2 and c (cartography analysis by precedent).

The version of the solution found with the help of information modeling environment, is referred to as *information model*. The model gives the description of the phenomenon or process under consideration by means of computer realized formalisms (knowledge and data base and algorithms). It includes the following elements:

- *Digital model* of researched property,
- *The rule* of calculation the model from the features,
- *Initial data, raster features* and *chains of operators* that are necessary and sufficient for the result.

Information models can be used for the solution of other similar problems. As new data and new expert knowledge are gotten, forecast quality of information model is improved.

4 GeoNet system

The Java-based client of the GeoNet environment is accessible on URLs http://www.iitp.ru/projects/geo and http://ta-www.jrc.it/gitis/index.html. Applet interactively supports the following functions of space-time data processing and analysis:

(i) Cartographic representation of raster, vector and point data:

- Composition and representation of maps, consisting from several raster, vector and point layers,
- Changing the sizes and scale of a map with and without interpolation,
- Reading values for all raster layers of a map in arbitrary points,
- Visualization of crossection by the user defind profile,
- Compiling similarity with precedent for arbitrary set of raster layers.

(ii) Data transformation:

- Shadow modelling,

- Calculation of raster fields with the help of user defined algebraic and logical operations performed on a set of initial rasters,
- Formation of learning and testing sample sets of precedents by means of selecting the individual points and/or specifying the precedents as all points of user defined polygons.

(iii) Plausible reasoning of forecast maps on a complex of raster maps [7, 8]:

- Reasoning in the method of similarity on a precedent set,
- Reasoning in the method of similarity on expert knowledge,
- Reasoning in the method of membership function for two classes,
- Reasoning in the method of nonparametric regression,
- Reasoning in the method of certainty function for classification on monotone feature space.

The maps displayed on Fig. 2 represent the following raster features: aeromagnetic anomaly, heat flow and topography.

5 Conclusion

The process of creating the information society is characterized by increased stream of geographically related space-time information. It invites the integration of geoinformation and communication technologies. GeoNet described in this paper is an approach to create such integrated intelligent technology that enables a researcher to complexly analyze

space-time properties of geological environment. The paper contains the description of the basic architecture of the system and its first version. The main functions realized in the system are the following: formalization of the problem, formalization of the information space of the problem, preprocessing, feature generation, generation of the problem solution, analysis and argumentation of the solution.

GeoNet is designed for various users with various queries. It results in necessity that information environment should be virtual, and its configuration must be adaptable to the demands of a user, so that it can solve both problems of acquaintance with the complex of geodata of a region of interest and research and practical problems of geoinformatics.

The easiest tasks of a user require analysis and comparison of large volume of assorted data. These data include the catalogues of points, polygons, polylines, time series measured in irregular region points, 2D and 3D raster data. Analysis of this information requires powerful tools of data processing, including space-time filtration, algebraic and logical operations on rasters, statistical analysis of point and raster data, cognitive cartography and the analysis on precedent. More complex queries are those that are related to narrow problem-oriented applications. Here decision making support is required. These tasks queries intelligent tools of data processing and analysis, such as problem-oriented methods of obtaining descriptive features, methods of representation of expert knowledge in the form of fuzzy logic constructions, methods of a plausible reasoning on space-time data, methods of hypotheses generation, methods of an explanation and confirmation.

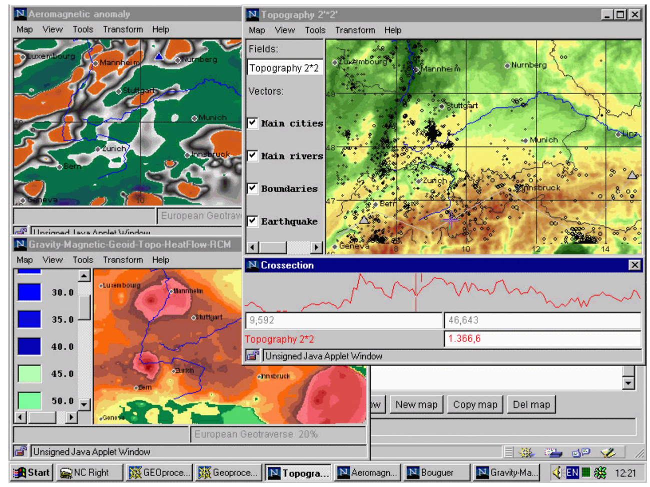


Figure 2 Working in GeoNet environment

The use of suggested technology in geological and geophisical research is only one field of application of network technologies for complex space-time data processing and analysis. Similar problems occur in different disciplines, such as economy, sociology epidemiology. Hopefully the described technology can applied in these areas, too.

The work is partially supported by the project ASPELEA granted by EU Inco-Copernicus Project contract \# ERBIC 15CT970200.

References

1. Proceedings of the 3-rd EC-GIS Workshop, Belgium - 25-27 June, 1997, edited by R.J. Peckham, 1997, 184.
2. Guidelines for best practice in user interface for GIS. BEST GIS, Esprit/Essi Project no. 21580, 104.
3. Gitis V.G. GIS technology for the design of computer-based models in seismic hazard assessment.- Geographical Information Systems in Assessing Natural Hazards, A.Carrara and F.Guzzetti (eds),1995, Kluver Academic Publishers, 219-233.
4. Gitis V.G., Jurkov E.F, Osher B.V., Pirogov S.A., Ponomarev A.V., Sobolev G.A. A system for analysis of geological catastrophe precursors.- Journal of Earthquake Prediction Research 3, 1994, 540-555.
5. A. V. Ponomarev, G. A. Sobolev, V. G. Gitis, Zhang Zchaocheng, Wang Guixuan, and Qin Xinxi: Application of GEOTIME Computer Environment to Space-Time Modelling of Earthquake Preparation Processes.- http://eos.wdcb.rssi.ru/journals/papers/9701/pap05

6. Gitis V., Vainchtok A., Koff G., Karagodina M. Geoinformation technology for seismic risk and losses assessment. - Applied geoecology, emergency situations, landuse cadastre and monitoring, no 2, Moscow, 1997, 38-42.

7. Gitis V., Jurkov E., Osher B., Pirogov S., Vainchtok A. Information technology for forecasting geological processes and phenomena. - Artificial Intelligence in Engineering 11 (1997), 41-48

8. Fukunaga K. Introduction to statistical pattern recognition. - ACADEMIC PRESS, New York and London, 1972, 321.