Use of fuzzy sets in modeling of GIS objects

Yu N Mironova

Elabuga Institute of Kazan Federal University, Kazanskaya, 89, Elabuga, Russia

E-mail: mironovajn@mail.ru

Abstract. The paper discusses modeling and methods of data visualization in geographic information systems. Information processing in Geoinformatics is based on the use of models. Therefore, geoinformation modeling is a key in the chain of GEODATA processing. When solving problems, using geographic information systems often requires submission of the approximate or insufficient reliable information about the map features in the GIS database. Heterogeneous data of different origin and accuracy have some degree of uncertainty. In addition, not all information is accurate: already during the initial measurements, poorly defined terms and attributes (e.g., "soil, well-drained") are used. Therefore, there are necessary methods for working with uncertain requirements, classes, boundaries. The author proposes using spatial information fuzzy sets. In terms of a characteristic function, a fuzzy set is a natural generalization of ordinary sets, when one rejects the binary nature of this feature and assumes that it can take any value in the interval.

1. Introduction

This paper discusses modeling and methods of data visualization in geographic information systems. Spatial data in geographic information systems are stored in the form of digital models; for their perception by a person, there is a need for data visualization. Usually this refers to their cartographic visualization. The object is described digital models of real spatial objects.

The author used materials from open sources: textbooks, articles on Geoinformatics, and also specialized sites. Such methods as analysis and comparison of data from different sources were used.

Problems of virtual modeling in Geoinformatics have been studied by many authors. So, if we consider the recent work in the field of Geoinformatics, the problems of virtual modeling are described in the work of Kapralov E.G., Koshkareva A.V., Tikunova V.S., etc. [1, 2, 3], Bulgakova S.V. [4]. Terrain modeling in geographic information systems was considered by Mironova Yu.N. [5]. There are materials on this subject on the Internet, e.g., on the Internet website of KB "Panorama" [6].

Thus, the problem of virtual modelling is currently relevant.

2. The main part (results of research and their discussion)

Information processing in Geoinformatics is based on the use of models. Therefore, geoinformation modeling is a key in the chain of GEODATA processing.

Modeling, as a method of scientific cognition, is a model and an action model [4].

Geoinformation modeling has several types: this is a simulation using digital models; spatial data models; GIS; using GEODATA and geo-information.
Geoinformation modeling is not only a means of displaying the phenomena and processes of the real world, but also an objective practical criterion for testing the truth of knowledge; it creates a new information models and information resources in the end.

When modeling, the original object is replaced by another object called the model. The model includes many parameters related to each other. Some parameters are to be determined based on measurements of the original object and are considered as a set of known values. Another part of the parameters is determined on the basis of calculations using the known parameters.

When considering real objects, mathematical methods are frequently used. Basing on the study of the characteristics of the object mathematical model, one receives some parameters. Changing these settings arbitrarily, one gets a change of the modeled object. Comparing the obtained results with the real ones, it is possible to improve the model and to predict further changes in the facility.

For GIS objects, it is convenient to visualize these parameters, that is, to show graphically on the map or on the 3D virtual terrain model. For example, one can show the changing weather conditions, sea currents, etc., obtained with the help of mathematical models.

It is possible to analyse developments over a longer period (e.g. a year or a few years), to clarify the corresponding mathematical models, and to visually show these changes (in the form of a video or series of graphic images).

This form of displaying information is more familiar to man than the manipulation of numbers.

Modeling makes it easier to reconstruct the processes of interaction between the real object and the external environment and to identify the criteria for the optimization of this interaction.

The geographic information modeling is the reliance on spatial relationships [4].

When rendering symbolic use of geoinformation modelling, the models become significant formations of any kind: card [3]; diagrams; graphs; drawings; formulas; graphs; symbols, etc.

In the study of phenomena or processes, it is preferable to use mathematical modeling. A mathematical model is a set of formal descriptions (formulas, equations, inequalities, logical conditions), which represent an actual process that changes the object state depending on various external and internal factors. The geographic information of mathematical modeling is the use of topology [7] and spatial data. In the study of spatial objects, digital simulation is widely used. In computer science and Geoinformatics [1, 2, 6], digital modeling is the realization of the potential of mathematical methods and software for modeling objects.

In the broadest sense, the digital model (TSM) (digital model DM) is a discrete model of information generated for computer processing.

In a narrow sense, the digital model is a discrete feature model, in which one of the mandatory parameters are: coordinates, sizes, dimensions, coordinate accuracy, scale, etc. This model is intended for processing in the information or geoinformation technology.

For the full human perception of spatial data, they must be expressed in the form of two-dimensional or three-dimensional models. In addition, animated models, three-dimensional dynamic models, etc. are often used. They are often used for practical purposes. The aim of the study is an overview of such models and their application in practice.

Methodology and methods of calculation become more complicated with increasing the reality of models and require the use of ever more sophisticated software for virtual modeling and computer resources.

2.1. The fuzziness of data in GIS modeling

When solving problems using geographic information systems (GIS), it is often necessary to represent approximate or insufficient reliable information about the map features in the GIS database.

Geospatial information arises mainly because of the need to solve geographical problems, many of which have social, economic and environmental value.

GIS, combined with remote sensing of the earth's surface from satellites, are advanced technologies to manage large amount of information relating to the compilation of the maps, to modeling and
decision-making. Geo-information technologies offer a wide range of functions, such as search, display, processing, decision support [4].

Methods for problem solving in GIS are not limited by geography. However, GIS requires a research orientation to support the technical development and application. Scientific orientation ensures accuracy in research, the application of common definitions to ensure accurate interpretations. However, the rapid expansion of GIS does not guarantee the accurate ensuring of the principles of collecting and using geospatial information. Do not confuse concepts such as geographical information and spatial information.

Geographical information differs from other types of information, their correlation with a certain place. In other words, it consists of facts or knowledge gained by researching, exploring special places on earth. Spatial information refers to information related to multidimensionality, which suggests a more General content than geography.

Geographical concept, formalized and implemented in GIS, reveals a more complex geographical shapes, models and processes. These can include topological relationships of adjacency, connectivity and intersection; the concept of simple and complex geographic features, etc.; the establishment of geographical concepts contributes to the development of GIS, helping to improve the scientific tools used to represent, study, modeling and understanding of objects.

Unfortunately, geographical data are often analyzed and associated with considerable uncertainty. Uncertainty is present throughout the process of geographic abstraction: from data acquisition to their use.

Geographic reality needs to be inevitable abstraction, simplification and fragmentation for analysis and decision-making. The process of abstraction and generation of real forms of geographic data are defined as data modeling. This process gives the conceptual model of the real world. The inevitable contradiction between the real world and the model is the inaccuracy and uncertainty, which can result in a wrong decision.

Modeling is complex interaction models, instrumental factors, and the acquisition of unproven components of each model of the data is also subject to multiple uncertainties. Depending on the level of the analyst and complexity of the instrumentation, the acquisition of data will have different levels of error. Although advances in technology, such as GPS, greatly improved the quality of spatial data, it is impossible to completely eliminate errors in their acquisition.

A large number of sources of performance data as well as GIS software of different standards and formats are often used. Here errors are found in the processing of geographical information (for example, transfer of the vector data structure to raster).

Heterogeneous data of different origin and accuracy have some degree of uncertainty. Uncertainty can be considered as a measure of differences between data and measurement, which a standard user assigns. In other words, uncertainty is a measure of the difference between the current contents database and the contents, which would be received by the user when accurate study of reality. Therefore, an increasing need for information with uncertainty-based GIS.

The geographical data recorded in a GIS, are the representation of some phenomena. For example, it is impossible to determine the true class of the soil, if the definition of classes of soil are mostly vague or inaccurate.

Therefore, the concept of "uncertainty" may be a more realistic view of geographical measurement than the concept of "error".

Another view of the GIS is based on errors: they should be presented as an integral part of human knowledge and understanding geographical reality, so the information is prepared in advance and used with known uncertainty estimates.

The imprecision and uncertainty exist in many geographical applications, reflecting the geographical characteristic of the world, such as complexity and subjectivity. The use of the term "uncertainty" indicates that there is some randomness of geographic information. In addition, uncertainty may result from lack of information.

The vagueness of objects is usually due to one or several reasons [4, 8], among which:
• the possible inaccuracy of information arising due to errors of sensors or due to inaccuracies or the inability of measurements of parameters of objects. The presence of this type of carelessness causes inaccuracy in the assignment of variables in the models, initial and boundary conditions;

• the inaccuracy of parameters is called the characteristic features of related objects;

• imprecision in the decision-making process, due to the fact that there are clear (explicit) objectives, models and corresponding solutions, does not exactly describe the system objects, and leads to quite large errors when making decisions because of ignoring some important factors;

• ambiguity is due to the impact of environmental factors.

The fuzziness caused by the first reason may be inaccurate measurement of the distance between service objects. For example, when measuring distance equal to several kilometers, the possible error of the measurement is equal to several centimeters or meters. When measuring the distance between complex objects, it is necessary to determine between which points of the objects one needs to measure (when measuring the distance between the two cities, districts, stations, etc.). The impossibility of measuring the parameters associated with the fact that some features have numerical values and are of good quality (quality of the road is impossible to measure, and can only be assessed as good, poor, very poor, etc.).

The inaccuracy of parameters, called the characteristic features of the objects, arises in cases when the objects themselves change their parameters (e.g., naturally they change the size of the reservoir) or their boundaries.

Imprecision in the decision-making process occurs in cases when the data about the goals and some of the parameters are not clearly defined or allow for variation in any range. It is advisable to describe the problem model in a fuzzy form. This will lead to more adequate description of reality and allow one to find a solution.

Ambiguity, due to the impact of the external factors, lies in the fact that the parameters of the objects or their relationships affect unexpected factors that change their size. For example, the travel time between two objects depends on the number of traffic lights and their status at the time of travel, weather conditions (speed in rain, fog and snow reduced compared to clear weather), the number of vehicles on the road, etc.

It should be noted that the formulation in the fuzzy form greatly reduces the possibility of getting incompatible solutions in the calculation and optimization.

The main drawback of deterministic models is the lack of effective methods for comparing different possible models for the purpose of the model, its accuracy and adequacy of the assumptions, on which it is based. The construction of the models in the framework of a fuzzy approach allows one to compare models. There is an opportunity of formalization of imprecise knowledge about the subject, entered into the model information incompleteness.

Let us consider what types fuzzy data are known at the moment, what operations one can perform on them and how to represent them in GIS databases.

2.2. The use of fuzzy set theory in Geoinformatics

Real world or real events can be defined either precisely or by using model representations (descriptions, images, mathematical, information model). The model can not match the original due to the large number of uncertainties, complexity, unclear definition of the real world. For treatment of uncertainties, sometimes it is necessary to use the laws of statistics and concepts of probability and reliability.

Geospatial objects in models were typically defined earlier in the properties. Applying the rules of logic and mathematics, the mathematical model operates on these objects, dividing them into two groups: satisfy/not satisfy; the selected rule; a "Yes" or "no" - without middle ground. These provisions do not allow overlap of classes, partial truth, partial membership.

In theory, multivalued logic, an ambiguity is defined as a type of uncertainty characteristics of classes that cannot have or do not have clearly defined boundaries. Blurring is often accompanied by complexity that is characteristic of geospatial objects.
However, the geospatial phenomena are described by many interrelated attributes, and their study is necessary to consider their attributes in the relationship. Thus, this interaction is unpredictable, so there is a need for applying the concept of fuzziness to geospatial information.

In geographical studies, GIS users usually have a clear idea of their purpose (for example, valuation of land or planning is determined uniquely). But they are often not sure where exactly should be the boundary between the classes of individual types of objects, which is inexact (or fuzzy). Identification of requirements should then translated in terms of the spatial objects in the database. In addition, not all information is accurate: already the initial measurements used poorly defined terms and attributes (e.g., "soil, well-drained"). Therefore, there are necessary methods for working with uncertain requirements, classes, boundaries [8].

The concept of fuzziness does not replace abstract concepts based on Boolean logic. Fuzziness is not a probabilistic attribute, which defines the membership of an object to a given class according to the statistical data using the well-known functions. More likely, the vagueness allows for the possibility of belonging, the valuation of which is based on subjective, intuitive or expert knowledge, but may correspond to a clearly specified uncertainty, for example, the possible measurement error of a certain size.

Standard, or crisp, sets allowing only binary membership is either 0 or 1. In terms of fuzzy sets, a degree of membership in a class is expressed in terms of the scale, which can continuously vary from 0 to 1.

In many cases, clear boundary classes are defined by their attributes in two ways:

1) construction of a discriminant function (separating) using expert knowledge, some regularity requirements; it is only necessary to set the upper and lower bounds of the class;

2) numerical taxonomy, for example, clustering.

Such path is also applicable for fuzzy sets to determine the boundaries of building discriminant functions $MF(z)$, called "membership functions".

For Boolean utensils such function has the form:

$$MF(z) = 1 \text{if } b_1 \leq z \leq b_2 \text{and } MF(z) = 0 \text{if } b_0 \text{ or } z > b_2,$$

where $z$ – set dimensions, which belong to a given class to determine; $b_1$ and $b_2$ - measurement, accurately defining the border of the class in the attribute space.

For fuzzy sets, membership functions are chosen to give the degree of belonging of the measure set in the center equal to 1, falling some way to fuzzy the border and taking a value of 0 outside the boundary. The point, at which $MF = 0,5$, is called the transition point.

The simplest $MF$ is a linear function that is specified by the pair of sloping lines intersecting at $MF = 1$ at point $c$, located in the center of set, and when $MF = 0,5$ crossing its border. The slope of the lines specifies the width of the fuzzy transition zone. The area between oblique lines and outside of the Boolean rectangle is called a zone of partial truth.

Many decisions depend on the accumulation of knowledge and experience of experts. The value of geospatial information is increasing especially with the use of software based on the technologies and methods of artificial intelligence and expert systems. They are intended for storage and use of special knowledge, there is connection between final and intermediate conclusions. This is especially important when dealing with fuzzy sets and fuzzy knowledge.

It is necessary to realize rationally such system by creating a special database of reference solution (test information, properly defined classes) and programs for the evaluation and comparison of them with the results (classes) based on, for example, compatibility matrix.

2.3. Fuzzy sets and their properties

The approach to formalizing the concept of fuzzy sets is the generalization of the concept of belonging. Let $U$ be a so-called universal set, which is formed by all the individual sets included in this class of problems (for example, the set of all integers). The characteristic function of set $A \phi U$ is function $\mu_A$, a value that indicates whether $x \in U$ is element of set $A$:
In terms of a characteristic function, a fuzzy set is a natural generalization of ordinary sets, when one rejects the binary nature of this feature and assumes that it can take any value in interval $[0,1]$. In fuzzy set theory, the characteristic function is called a membership function and its value $\mu_A(x)$ - the degree of membership of element $x$ to fuzzy set $A$. Moreover, fuzzy set $A$ is the set of pairs:

$$ A = \{ < x, \mu_A(x) > | x \in U \} $$

where $\mu_A$ – the membership function, that is $\mu_A: U \rightarrow [0,1]$.

For any fuzzy set operators $F = \min$ and $G = \max$ [see 8] are the only operators at the intersection and Union of the following properties:

1) Commutativity

$$ F(\mu_A, \mu_B) = F(\mu_B, \mu_A), G(\mu_A, \mu_B) = G(\mu_B, \mu_A) $$

2) Associativity

$$ F(\mu_A, F(\mu_B, \mu_C)) = F(F(\mu_A, \mu_B), \mu_C), G(\mu_A, G(\mu_B, \mu_C)) = G(G(\mu_A, \mu_B), \mu_C) $$

3) Distributivity

$$ F(\mu_A, G(\mu_B, \mu_C)) = G(F(\mu_A, \mu_B), F(\mu_A, \mu_C)), G(\mu_A, F(\mu_B, \mu_C)) = F(G(\mu_A, \mu_B), G(\mu_A, \mu_C)) $$

4) Monotony

$$ \mu_A \leq \mu_C, \mu_B \leq \mu_D, \text{ then } F(\mu_A, \mu_B) \leq F(\mu_C, \mu_D), G(\mu_A, \mu_B) \leq G(\mu_C, \mu_D) $$

$$ \mu_A \leq \mu_B, \text{ then } F(\mu_A, \mu_A) < F(\mu_B, \mu_B), G(\mu_A, \mu_A) < G(\mu_B, \mu_B), \quad F(1,1) = 1, G(0,0) = 0 $$

$$ F(\mu_A, \mu_B) \leq \min \{ \mu_A, \mu_B \}, G(\mu_A, \mu_B) \geq \max \{ \mu_A, \mu_B \} $$

2.4. The main practical application of virtual modeling

The main practical application of virtual modeling [2]:

1) Cultural-historical model, realistic restoring various historical eras, events, landscapes (it can be used in museums, schools, universities).

2) Training of pilots of different aircraft control and orientation in unfamiliar terrain (this is especially important for pilots of small aircraft flying in a mountainous area).

3) Strategic planning of military operations.

4) Planning of major economic projects.

5) Marketing and promotional activities.

The author is currently focused on increasing the efficiency of digital information about the area in automated command and control systems, navigation systems, command and defeat in the targeting and in simulators [9].

3. Conclusion

The work considers the modeling and visualization techniques of spatial data in geographic information systems and their application in practice. The author also analyzed the possibility of applying fuzzy logic in GIS modelling. One can draw the following conclusions: geographic information systems and visualization of data in them are rapidly evolving, due to the rapid development of computer technology, the emergence of new methods of data collection, development of the methodological base. In this regard, there are new ways to use them in practical activities, in particular, in elimination of emergency situations in real time, in the military field, but also in daily activities of organizations and citizens.

References

[1] Kapralov E G 2010 Geoinformatika: v 2 kn. Kn. 2: uchebnik dlja stud. vyssh. ucheb. zavedenij (Moskow: Izdatel'skij centr «Akademija»)

[2] Kapralov E G 2010 Geoinformatika: v 2 kn. Kn. 1: uchebnik dlja stud. vyssh. ucheb. Zavedenij (Moskow: Izdatel'skij centr «Akademija»)
[3] Kapralov E G 2013 Sbornik zadach i uprazhnenij po geoinformatike: ucheb. posobie dlja stud. vyssh. ucheb. Zavedenij (Moskow: Izdatel’skij centr «Akademija»)

[4] Bulgakov S V 2013 Osnovy geoinformacionnogo modelirovaniya Izvestija vysshih uchebnych zavedenij. Geodezija i ajerofotosiemka 3 77-80

[5] Mironova Yu N 2015 Matematicheskie aspekty geoinformatiki Internet-zhurnal «NAUKOVEDENIE» 7 - 5

[6] Internet-resurs KB «Panorama» http://www.gisinfo.ru

[7] Skvorcov A V, Sarychev D S 2002 Tehnologija postroenija i analiza topologicheskijh struktur dlja geoinformacionnyh sistem i sistem avtomatizirovannogo proektirovaniya Vestnik Tomskogo gosudarstvennogo universiteta 275 60-63

[8] Jahjaeva G Je 2006 Nechetkie mnozhestva i nejronnye seti: Uchebnoe posobie. (Moscow: BINOM)

[9] Mironova Yu N 2016 Novye metody virtual'nogo modelirovaniya v geoinformacionnyh tehnologijah Internet-zhurnal «NAUKOVEDENIE» 8 - 5