Formation of a data network in a geo-information system for forest inventory

M Kalambet¹, V Mikheev², E Istomin², N Yagotinceva² and M Vagizov³

¹Joint-stock company "Masshtab research institute", Kantemirovskaya Street 5, letter A, St. Petersburg 194100, Russian Federation
²Department of applied information science, Russian state hydrometeorological university, Voronezhskaya Street 79, St. Petersburg 192007, Russian Federation
³Department of information systems and technologies, Institute of forest and nature management, State Saint-Petersburg forest technical university named after S. M. Kirov, Institutskii Lane 5, St. Petersburg 194021, Russian Federation

*Corresponding e-mail: yagotintceva@yandex.ru

Abstract. This article considers the application of intelligent geographic information system in forestry. A distinctive feature of the proposed system is the use of data processing module included in the GIS. Ways of increasing the efficiency of network resources are also considered. Recommendations have been formulated, which will enhance the quality of service, transmission speed, efficiency and scalability of routing protocols, cost and signal quality in GIS.

1. Introduction

GIS technology has emerged with the advent of second-generation high-speed computers and data storage systems. This technology is applicable wherever spatially distributed information needs to be processed. A promising area of GIS application is the provision of information to management systems of objects in geospace, control and support the decision-making. One of the promising technologies of GIS processing is the technology of intelligent data analysis integrated into functional environment. The main components are implemented on the basis of individual software-algorithmic solutions that do not require the direct participation of the decision maker in the analytical process and final result. It is possible to apply these methods when processing data. Its volume exceeds hundreds of terabytes and is constantly increasing over time up to ultra-large volumes of information [1].

Data mining techniques are flexible and applicable to a variety of human activities. Forestry and forest management are being no exception. However, one cannot speak only about databases and the information contained in them. The scope of forestry tasks also includes the transfer of analyzed data. The main challenge in this context, apart from the technical implementation, is to improve the quality of the forest information.

The forest assessment is to provide information for the decision makers at different levels of government. The challenge of the forestry sector is the ever-changing forest information. There are four main cycles in the collection of forest information, presented in figure 1.
In terms of the source of collection, transmission, processing and interpretation the information in each of the above stages can be generated by both decision makers and devices. In this article, the data transfer stage without explicit losses will be considered. The solution of problems of qualitative and quantitative forest accounting depends on the speed and completeness of big data transfer. The development of routing systems in the data network, in terms of the formation of non-coordinate (attributive) characteristics, involves solving the need to operate with large volumes of heterogeneous geodata [2].

Spatial data, imagery and metadata are used as data to be handled by intelligent geographic information system. The format of the data determines the way geospatial information is processed. Symbolic data contained in databases are processed due to laid down rules and queries to this data, processing of graphic information from images is more complicated, due to the structure of the data [3]. From the procedural point of view the images of forest-covered land must first be decoded into comprehensible sets of numerical data. For this purpose the image is broken down into constituent elements. In this case, conventional classification by systematic analysis methods does not always yield a perceptible result. In fact, the pixel-form graphic data is a two-dimensional data set, but it is represented initially as a single data set in the RGB (Red, Green and Blue) colour model.

2. Methods and Materials

2.1. Conceptual representation of GIS data transmission

When designing the GIS, spatial geodata (coordinates and attributes) must be included. It is the base of GIS. Spatial data units: point, segment, polygon, data from the last inventory (composition, predominant species, age, quality class, fullness, stock, forest type, area of the stand). Speaking about image data representation we mean graphical data (multi-scale images and fragments, exported images of covered forests in web-mapping services). They can be presented in raster and vector format [4].

Characteristics include date and time of capture, location, camera used, altitude, shooting direction, scale, etc. These characteristics will help to link the obtained image to the existing map. They also help to trace the dynamics of changes when comparing two remote sensing materials of different dates of surveying the same area.

To ensure that all levels of information processing in an intelligent GIS are workable, the key aspects of the system can be highlighted:
- a large amount of heterogeneous data,
- improved service quality,
- data rates,
- high efficiency and scalability of routing protocols,
- signal quality in GIS,
- hardware.

The hardware is an integral part of the GIS subsystem. Therefore, an effective GIS synthesis, to support decision-making for route finding in the data network, will also depend on the correct configuration, the technical component of the computer on which the system will be based.

The synthesis of data transmission hardware in the intelligent GIS data transmission network refers to the second class of problems (figure 2).
Figure 2. Classification of tasks for improving the efficiency of information systems.

For this set of problems a selection of options is made, which achieve the set goal with a minimum of resources. In the process of research it is necessary to design the structure of the future GIS platform. It should be achieved by methodological approach to the analysis and processing of forest information. The structural level of the system implementation is presented in figure 3.

Figure 3. Structural level of system presentation.

2.2. Mathematical model for calculating GIS data transmission

An intelligent GIS features layer-by-layer data arrangement. The basic element in any GIS is data layers. In IGIS, however, the layers can be the results of database queries, forecasting and modelling. These layers reflect the current situation and changes. They may be called dynamic map layers. Implementation of such layers can be done by using application programming tool - API-maps [5].

Cross-platform in GIS contributes to the increase in the amount of data due to the growing hardware interaction between the user and the system. The goal of the IGIS is to solve the problems of processing large volumes of information. The main tasks include quantitative and qualitative
accounting of the forest. In order to solve the IGIS tasks it is necessary to consider certain functions presented in table 1.

**Table 1. The main types of data processed by IGIS.**

| Standard functions          | Intelligent data processing functions | File type | Data source                          |
|----------------------------|---------------------------------------|-----------|--------------------------------------|
| Downloading forest maps    | Highlighting reference objects on the map | Layers    | Forest district maps, forest plantation plans |
| Map processing             | Creating dynamic layers               | Layers    | Earth remote sensing data             |
| Tracking illegal logging   | Predicting vulnerabilities to possible illegal activities | Dynamic layers | Remote sensing data, forestry maps |
| Adding and editing data    | Data processing, analysis and display to the user | Graphs    | Databases                            |

The conceptual representation of intelligent GIS decision support allows to propose the structural and functional model. The proposed model details the process of interaction between the intelligent decision-making system and the hardware part of the GIS. The transport basis of the distributed GIS is a computer network with packet routing (data in the network). The technology makes it possible to transmit data simultaneously between the all geographically distributed interacting devices [6].

The task of routing is to select a route for transmission from a sender to a receiver. It makes sense in networks where optimal and acceptable route selection is necessary or possible. The selection of routes in GIS communication nodes is done according to the routing algorithm (method). The routing algorithm is the rule for assigning an output link of a given GIS communication node to transmit a packet based on the information contained in the packet header (sender and receiver addresses), the information about the load on that node (packet queue length) and possibly the information about the entire GIS.

The routing problem in the networks is solved under the assumption that the shortest route for providing packet transmission in the shortest time depends on the network topology, the capacity of the communication lines, and the load on the communication lines (figure 4).

![Route layout in a typical network.](image.png)
The network topology changes as a result of node and link failures and partly with the development of the GIS (connection of new nodes and links). The capacity of communication lines is determined by the type of transmission medium and depends on the noise level and the parameters of the equipment serving the lines. The most dynamic factor is the load on the links, which changes quite rapidly and in a direction that is difficult to predict.

In order to select the best route, each communication node must have information about the state of the GIS as a whole - all other nodes and links. Data about the current network topology and link capacity are provided to the nodes without difficulty. However, there is no way to accurately predict the network loading. Therefore, when solving the routing problem, load data may be used, which lags (due to finite information transmission rate) with respect to the time of deciding on the direction of packet transmission. Consequently, in all cases, routing algorithms are executed under the uncertainty of the current and future states of the GIS.

The operation of an intelligent routing system is based on the use of:
- a method for multi-criteria optimization of routing parameters in geographically distributed networks based on a genetic algorithm;
- bottleneck search method in geographically distributed networks based on the Laplace-Stiltjes transform.

A genetic algorithm searches for the best solution in the search space, which, under the influence of evolution mechanisms, changes in the direction of "improvement" of the contained solutions. The result of such evolution should be a search space containing the best (or acceptable) solutions, which are discovered by the algorithm. Genetic algorithms differ from traditional optimization methods in several basic elements, in particular:
- the coded form of the parameters of the optimization problem is processed, not the values of these parameters;
- the search for a solution is not based on a single point, but on some population of points;
- only the target function is used, not its derivatives or other additional information;
- probabilistic rather than deterministic selection rules are applied.

These properties determine the superiority of GA (Generic Algorithm) over other structural optimization methods. In GA terminology, the solution space is interpreted as a population 

\[ P = \{ A_1, \ldots, A_k, \ldots, A_N \} \]

and each solution from that space is like a chromosome (or an individual) \( A_k \), represented as a string of characters called genes:

\[ A_k = \{ a_1, \ldots, a_j, \ldots, a_L \} \]  \hspace{1cm} (1)

This is done by coding independent chromosomes in either binary or floating-point format. Then the genome in this chromosome will be one bit. With each gene \( a_j \) the set of its possible values - alleles ("0" or "1"). On the set of solutions (chromosomes) the target function (TF) is defined - the utility function \( F = (A_k) - \) decision performance \( A_k \). As a utility function \( F(A_k) \) is assumed to be the state in which for each task with performance requirements there exists a resource with current performance for that chromosome \( A_k \). The solution to the problem is to calculate the allocation route for available resources of the available tasks using the GA \[7\]. The GA decision-making process is represented as a flowchart in figure 5.

The sequence of the GA, which forms the plan, consists of the following steps.

1. Creating an initial population \( P_0 \) consists of performing an initialization operation for each individual of its \( N_0 \) chromosomes, i.e. randomly allocating resources to tasks. One chromosome initialization algorithm: forming a random list of resources; forming a random list of tasks; selecting the next task and resource from the list, if it is not empty, otherwise terminating the algorithm; matching the selected task to the resource.

2. Calculating the quality of the population. This step is important for determining the value of the GA stopping criterion. The utility is calculated \( F = (A_k) \) of each chromosome and the utility of the whole population. If the whole population is viable, then a distribution route is formed from it. If not, move onto step 3.
3. Selection of chromosomes for inbreeding. In this implementation of the algorithm, the method of proportional selection of chromosomes directly from the preceding population (in the first iteration - from the initial one) is chosen. Non-viable individuals are removed from the population and supplemented with new ones. A second population is formed at random, and then the next step is taken.

4. Execution of the genetic algorithm operators. Pairs of different chromosomes are formed from the obtained individuals, and a crossing-over operation is performed on each of these pairs with a given probability. If the offspring does not meet the constraints or if the utility of the offspring is less than the utility of the parent chromosome, the offspring is discarded and a new parent pair is sought for cross-over. If the offspring satisfies the constraints and the utility of the offspring is greater than the utility of the parent, the parents are discarded. And move onto step 2.

The most useful chromosome in the population, obtained before stopping, is the solution to the problem. Thus the GA computation process takes place in a procedural routing mechanism. The result is the selection of optimal points to construct a geodata route with minimum losses and delays. For a more accurate performance of this system, a bottleneck search using the Laplace-Stilts transform (LST) is performed before selection to estimate the network performance. Using the Laplace-Stiltsky transform of the distribution function $B(t)$is called a function $s$, defined as follows

$$\beta(s) = \int_0^\infty e^{-st} dB(t_i)$$

where, $s$ – LST parameter.

The result is the probabilistic meaning of the Laplace-Styles transform. The value of $e^{-st} B(t_i)$ is the probability of the complex event that a random variable will not exceed the value of $t_i$ (multiplier $B(t_i)$), also in the time it takes to $t_i$ no "catastrophe" will occur (the multiplier is $e^{-st}$). The parameter $s$ is regarded as the intensity of the "catastrophes". Integration over the whole range gives

$$\int_0^\infty e^{-st} dB(t_i) = \beta(s).$$

Figure 5. Scheme of a classical genetic algorithm.
Thus, the probabilistic meaning of the Laplace-Stiltsky transform is that it determines the probability that in time $t_i$ no single "catastrophe" having the greatest impact on the transmission delay will occur or will fail at all.

3. Conclusion
This paper considers a model for calculating intelligent GIS decision support for the route search in a data transmission network consisting of two levels. The model allows applying heuristic methods, concepts and modern tools to ensure the processing and transmission of packets in the data transmission network.

The use of genetic algorithm to solve the problem is determined by two main factors: speed and stability. The speed of a genetic algorithm is evaluated by the time required to execute the user-defined number of iterations. An additional quality of optimality and speed of finding a solution is achieved by pre-selection bottleneck search in geographically distributed networks based on the Laplace-Stiltjes transform.

References
[1] Malikova D M, Slashchev E S, Istomin E P, Vagizov M R and Kolbina O E 2020 Method for solving problems of the theory restrictions of infocommunication systems using linear equations with many unknowns. *IOP Conference Series: Earth and Environmental Science* vol 574 (1)
[2] Bykov F L 2020 Statistical correction of COSMO weather forecasts using neural networks [in Russian – Meteorology and hydrology] vol 3 pp 5-20
[3] Kvochkin, D.O., Ustyugov, V.A. 2017Mobile device of the automated forest inventory. *Journal of Industrial Pollution Control*. vol 331(1) pp 976-980
[4] Kalambet M, Yagotinceva N, Kolbina O, Yagotinceva T and Aksenova A 2020 Opportunities and prospects of IoT application in landscape architecture, design and information and communication technology. *IOP Conference Series: Earth and Environmental Science* vol 507 (1)
[5] Istomin E, Petrov Y, Stepanov S, Kolbina O and Sidorenko A 2019 About technology of risk management in forestry. *IOP Conference Series: Earth and Environmental Science* vol 316(1)
[6] Bogatyrev V A, Bogatyrev S V and Bogatyrev A V 2012 Of the computing systems with redistribution of queries [in Russian – Izvestia vuzov. Instrumentation.] No 55 vol 10, pp 53-56
[7] Shishkova N A 2017 Application of genetic algorithm to solve the traveling salesman problem [in Russian – education and culture] (Ivanovo: Olympus) No 4 (19), pp 6-7