UDC: 623.618:004.657 https://doi.org/10.15407/jai2022.02.078

DATA MINING AS A MEANS OF AUTOMATION OF THE DECISION-MAKING SUPPORT IN AUTOMATED TROOPS MANAGEMENT SYSTEMS

Horhulenko V.
Central Scientific Research Institute of the Armed Forces of Ukraine
Povitroflotskyi Avenue, 28, Kyiv, 03049
vladyslavhorhulenko@gmail.com
https://orcid.org/0000-0001-6382-5075

Abstract. The article examines the technology of data mining as an instrumental means of automating decision-making support in automated troop management systems (ATMS). During the study of the subject area, the problems of the decision-making support process in ATMSs have been analyzed, sources testifying to the relevance of the automation of this process and the current state of affairs are being considered. An option to increase the degree of automation of the decision-making support process is proposed for consideration due to the use of intelligent data analysis technology, which has been the main focus of the article. During the analysis of the latest research and publications on this topic, it has been found that the automation of the decision-making support process due to intelligent data analysis is considered promising and is studied, as a rule, within the framework of research on artificial intelligence and machine learning technologies. The main material of the study covers the generalized algorithm of automated decision-making support, an analogy of the functions of the intelligent decision support system with the tasks solved by intelligent data analysis is also made. Based on this, the clustering tasks have been described, search for associative rules, classification and regression in the context of their application for getting knowledge from data found in the automated troop management system, with examples of the results of their implementation. A more in-depth study of intellectual data analysis is defined as directions for further research and ways to automate the decision-making support process using this technology. Also, a promising direction of research is simulation modeling of an intellectualized decision support system of an automated troops control system.

Keywords: data mining, data analysis, big data, decision support system, automated troops management system.

Formulation of the problem in general

In the document defining the main directions of implementation of the military policy of Ukraine – Strategic defense bulletin of Ukraine, one of the problems of the functioning of the defense forces is the incompleteness of the creation of an automated system of operational (combat) control, communication, intelligence and surveillance, and the direction of achieving a promising model of the Armed Forces of Ukraine – application of ATMSs during the operation (combat operations) of the united grouping of troops (forces) [2].

In [3], the course of the Ministry of Defense of Ukraine on the development of the communication system, control automation and situational awareness is expressed, taking into account the practical experience of the Russian invasion of Ukraine. Taking into account the fact that the management of troops necessarily entails decision-making, accordingly, the automation of management is the automation of this decision-making.

As ATMS, as a rule, solve problems that do not have an unambiguous clear solution due to the multi-criterion of their evaluation, decision-making takes place in conditions of uncertainty [1].

Automation of decision-making in the ATMS should mean the provision of a decision maker (DM) by a set of optimal decisions and conclusions from the analysis of available data at all stages of troop management (in real time).

On the other hand, in the ATMS "Dzvin-AS", which is at the stage of research and design work and is intended to perform the functions of the Unified ATMS of the Armed Forces of Ukraine a complex of units, modules and services has been implemented. For example: a geographic information system (GIS), information and calculation tasks (ICT) with references to GIS, a service to create and editing the organizational structures of military units and groups for these ICTs. Also, there is a functional for creating troops management documents of various nature, providing access to them to other officials, thereby official correspondence has been realized. At the same time, in the composition of this ATMS, regardless of the results provided by the ICTs, there are no functional capabilities for providing the DM with optimal solutions obtained on the basis of forecasting and prediction in an automated mode.

It is possible to increase a level of automation of the decision-making support process in the ATMS by focusing on the analysis of data that can potentially contain useful information, thereby simulating the "learning" of the system. Moreover, in such a concept the larger the volumes of analized data, the higher the relevance of the decisions formed by the Decision Support System (DSS) for DM.

With the development of information technologies, a technology such as data mining (also known as knowledge discovery in databases) has gained considerable attention.

The main purpose of data mining is to identify previously unknown patterns in large arrays of information and to present them in a visual form for DM. Merely the consideration and presentation of data mining technology as a means of automating decision-making support in the ATMS this article will be focused on.

Analysis of recent research

The understanding of the importance and potential benefits of analyzing of information, which is entering the ATMSs and, in general, computerized systems arose at the end of the 20th century, first of all, among scientists from the United States. Research in the direction of data mining and related technologies (big data technology, artificial intelligence, etc.), as well as their application in various military information systems and, in particular, in ATMSs, is constantly being conducted.

For example, at the conference that took place in November 2020, scientists from NATO member states shared the results of research in the field of automation of management decision-making based on Big Data analysis and artificial intelligence [4].

It is also worth noting, in the context of this study, that scientists from the Dalian Naval Academy (China) developed the model of the maritime operations ATMS based on data mining technology, which describes the advantages of using the latter for processing large arrays of combat data, getting from them useful information (knowledge) and providing support to the DM in decision-making [5].

Purpose of the article

The purpose of the article is to examine the data mining technology as a means of automation decision-making support in ATMSs.

In order to achieve the purpose, the following partial research tasks have been decided to be solved:

- to analyze the classification of modern military (combat) DSSs according to the level of automation of their functions in order to focus on merely the type where the automation of decision-making takes place;
- to determine (find out) the order (algorithm) of decision-making support in such DSS;
- to examine the data mining as an instrumental means of automation decision-making support in ATMSs.

Presentation of the main research

Under the influence of the modern nature of military (combat) operations, which are characterized by dynamic, accelerated changes in the environment, the ability of the DM to make proper decisions in a timely manner at all stages of the operation has acquired a particularly important role. Invaluable help in solving this problem is provided by the DSS as a structural element of the ATMS. It is the DSS as a part of the ATMS that is the main tool for improving the validity of management decisions [1].

However, DSS does not have one stable definition. A collection of their variations is summarized and given in [6]. Also, from the point of view on the level of user functions automation, that are implemented in the decision-making process, it is proposed to adopt as the main classification of the DSSs according to the level of DM's participation in the ATMS's functioning. According to this feature, military (combat) DSSs can be divided into the following groups:

− informational («C2» – Command and Control);

− calculational and informational («C4» – Command, Control, Communications and Computers);

− intellectualized («C4ISP» – Command, Control, Communications, Computers and Intelligence Support Plan, the most close).

In order to consider DSS as intellectualized (IDSS), it is necessary to have in such system a set of tools for working with knowledge, and sometimes for solving partial tasks in the general cycle of decision-making support, some elements (algorithms) of artificial intelligence are also necessary [6].

Further in the article the automation of decision-making support will be considered as a phenomenon (process) merely in the IDSS-like systems.

The key difference of such DSSs is a capability to provide the DM with a set of optimal decisions (solutions) based on the knowledge that is formed using special data analysis algorithms. That, for obvious reasons, can be called a "real" decision-making support, in contrast to the usual functions of calculational and informational DSSs, which only provide results according to the options included into the system by the DM.

At this stage it is clear that the methodological basis for the automation of the decision-making support is conceptually laid precisely in the IDSS-like systems.

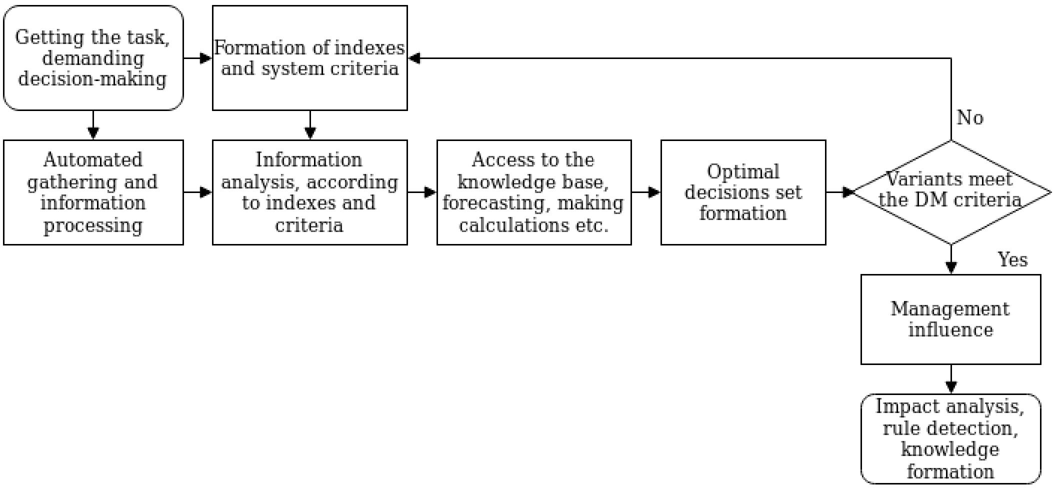

Fig. 1. Generalized scheme of the decision-making support algorithm of ATMS's IDSS

Researching and comparing of different decision-making support algorithms of ATMS's IDSS [3] – [7] gave possibility to form their generalized scheme (fig. 1).

Data mining function (purpose) is an extraction of previously unknown and non-trivial knowledge [8], which is represented by models. The types of models depend on the methods of their creation. The most widespread are: rules, decision trees, clusters and mathematical functions.

Methods of the data mining are useful to solve a set of problems (tasks) faced by the analyst (DM). These tasks can be divided by a result entity type [8]:

− descriptive tasks (clustering, search for associative rules);

− prediction and forecasting tasks (classification, regression).

The functions performed by IDSS can be logically divided into: analysis of the current situation conditions and decisions made in these conditions (filling the knowledge base); forecasting and providing a set of optimal solutions (decisions) according to situation conditions (applying to the knowledge base) [7].

These two functions of IDSS are correlated with the tasks of data mining, which are described above. Complete description of the data mining tasks are given in [8]. Further, a briefly consideration of the main ones will take place, in terms of their use for automation decision-making support in ATMS.

The task of clustering consist of division of a studied information set into groups of "similar" objects in terms of parameters, called clusters. Formally, the task of clustering should be described as follows: from a set of given data objects I, it is necessary to build a set of clusters C. Quality of the solution for this problem is determined by the number of correctly qualified data objects.

The set I will be defined as:

$$I = \{i_1, i_2, \dots, i_j, \dots, i_n\}, \qquad (1)$$

where i_j – the studied data object;
n – a total number of data objects.

Each of the data objects of the set I is characterized by a set of parameters:

$$i_j = \{x_1, x_2, \dots, x_h, \dots, x_m\}, \qquad (2)$$

where x_h – parameter of the data object i_j;
m – the total number of i_j data object parameters.

Accordingly, each parameter x_h can take values from the set:

$$x_h = \{v_h^1, v_h^2, \dots\}. \qquad (3)$$

The task of clustering is to form the set:

$$C = \{c_1, c_2, \dots, c_k, \dots, c_g\}, \qquad (4)$$

where g – number of created clusters;

c_k – the cluster containing objects of the same type from the set I:

$$c_k = \{i_j, i_p | i_j \in I, i_p \in I, d(i_j, i_p) < \sigma\}, (5)$$

where σ – is a is the value that determines the degree of approximation between data objects for adding them to the same cluster;
$d(i_j, i_p)$ – a measure of proximity between data objects.

For example, the following clusters can be derived from the test data set of high-precision weapon’s (HPW) using (Table 1):

− significantly distant – {{3, “Kh-47M2 Kinzhal”, 06:05, Vinnytsia, military}, {5, “Kh-55”, 21:25, Rivne, critical}};

− remote – {{2, “Kh-101”, 05:00, Kyiv, critical}, {1, “Kh-55”, 10:40, Odesa, critical}};

− close – {{4, OTR-21 “Tochka”, 14:30, Luhansk, civilian}}.

Information, which has been shown in the table below, was arranged randomly, only for data mining tasks and methods testing in military context with no accordance to the real combat events.

Table 1 The test data set of HPW’s using

Num..	Type of HPW	Time	State	Infrastructure
1	“Kh-55”	10:40 am	Odesa	critical
2	“Kh-101”	05:00 am	Poltava	critical
3	Kh-47M2 Kinzhal	06:05 am	Vinnytsia	military
4	OTR-21 “Tochka”	02:30 pm	Luhansk	civilian
5	“Kh-55”	09:25 pm	Rivne	critical

In this case, the clusters have been formed according to the distance of the affected objects from the zone of active hostilities (70 km). Cluster analysis of real volumes of data according to a specified parameter would allow, for example, to distinguish a cluster of the enemy's strategic goals, for appropriate decision-making.

The task of searching of associative rules is used to determine combinations of parameters in sets of data objects that are often found in large volumes of information. In this context, the shown above data objects are considered as transactions T. By D will be meant the set of transactions available for analysis.

There is the set of parameters

F = {significantly remote, military infrastructure}.

The set of transactions, which includes the set F will take the following form: D_F = {3, Kh-47M2 Kinzhal, 06:05 am, Vinnytsia, military}.

The ratio of the number of transactions in which the set F is included to the total number of transactions is called the support of the set F, and is denoted by $Supp(F)$:

$$Supp(F) = \frac{|D_F|}{|D|}, \quad (6)$$

For set F, the value of support will be 0.2, since it is only relevant for one transaction out of five. Thus, we form a rule: *"the enemy using of HPW on military infrastructure with significantly distant from the combat zone is no more than 20% of the total number"*. ATMS's DSS can use this rule for providing the DM with recommendations regarding, for example, changing the points of permanent deployment of military units located in the rear in those cases when it is possible.

The descriptive tasks of data mining have been described above, let's move on to the tasks of prediction and forecasting.

The essence of the tasks of classification and regression comes down to determining the dependent parameter of the analysis data object on the basis of independent ones. Their difference lies in the type of the output parameter: for classification it is expressed qualitatively (the listed type), for regression – quantitatively (a value from a set of real numbers).

Formula 2 will be expanded with a dependent variable y:

$$i_j = \{x_1, x_2, \dots, x_h, \dots, x_m, y\}. \quad (7)$$

Tasks of classification and regression, in the general algorithm of the ATMS's DSS (Fig. 1), make it possible to attribute events that may occur to defined classes, or to predict certain numerical indicators (obtain the value of the dependent parameter y based on the regularities between independent parameters x_h) on the basis of already analyzed data sets. For example, in the context of anti-aircraft missiles using by the enemy, it is possible to predict the probability of damage to an extremely valuable infrastructure object, as well as to determine the distribution of air defense means based on the conducted analysis.

Classification and regression results can be displayed by classification rules, decision trees and mathematical functions, formed according to special algorithms. Algorithms of classification and regression are qualitatively described in [8].

The implementation of data mining tasks, which are described above, as a part of the ATMS of the Armed Forces of Ukraine (as well as other military formations in other countries), will allow to reformate the system into one that independently learns from data. What will certainly result in more quickly decision-making and increasing of these decisions validity, on the basis of powerful opportunities provided by technology of data mining.

Conclusions and prospects for further research

In the study has been analyzed the current state of modern ATMSs, approaches to support decision-making, and classification of DSSs in these systems. As a result, it was found that, despite the wide list of components of the ATMS, the DM makes decisions on the basis of its own analysis, while the systems are oversupplied with data that potentially contain useful information that can be used to form rules, regularities and, thus, get hidden knowledge them. It is knowledge that is the key to automated decision-making support, because it is obtained from large sets of data (real facts). Conceptually, such a mechanism is embedded in the IDSS. The main attention during the research has been paid to the consideration of the data mining technology as an instrumental means of automation decision-making support, and as a result, main tasks of this technology have been described.These tasks allow to identify regularities in data sets for the formation of the optimal decision set for the DM in the IDSS-like system.

The following directions for further research emerge from the results of this study,such as: formation of the functional scheme of data mining process in the IDSS; determination of approaches to the formalization of a typical data model of ATMS for the data mining implementation; in-depth research in the direction of machine learning, the concept of which is based on the principle of self-improvement by the system without an intervention of a human analyst. Also, one of promising directions for further research is using of data mining tasks and method for the analysis of military (combat) simulation mathematical modeling results.

References

1. Decree of the President of Ukraine No. 473/2021. On the decision of the National Security and Defense Council of Ukraine dated August 20, 2021 "On the Strategic Defense Bulletin of Ukraine".

2. Communication, automation of management and provision of situational awareness are the subject of special attention of the Ministry of Defense. Army Inform [Electronic resource]. Access mode: https://armyinform.com.ua/2022/06/30/zvyazok-avtomatyzacziya-upravlinnya-ta-zabezpechennya-sytuaczijnoyi-obiznanosti-ye-predmetom-osoblyvoyi-uvagy-minoborony/
(Date of application: 15.07.2022).

3. Decision support systems: textbook and practical guide for undergraduate and graduate studies / Ed. V.G. Khalyna, G.V. Draft. - M.: Yurayt Publishing House, 2015. - 494 p.

4. NATO Decision-Making in the Age of Big Data and Artificle Intelligence. Edited by Sonia Lucarelli, Alessandro Marrone, Francesco Niccolo Moro. – Brussels: NATO HQ Brussels, 2021. – 100p.

5. Hong Wen Lin, Jian Chen, Yang Song. Warship Combat Decision Support System Based on Data Mining Technology: 3rd International Conference on Mechatronics and Intelligent Robotics. 2020. – 588 p.

6. Tykhanychev O.V. Theory and practice of decision support automation.-M.: Editus, 2018. - 76p.

7. Gerasimov B.M., Lokaziuk V.M., Oksiyuk O.G., Pomorova O.V. Intelligent decision support systems: Education. manual. - K.: View of Europe. University, 2007. – 335 p.

8. Analysis of data and processes: tutorial. manual / A.A. Barseghyan, M.S. Kupriyanov, I.I. Kholod, M.D. Tess, S.I. Elizarov. - 3rd ed. - St. Petersburg: BHV-Petersburg, 2009. - 512 p.

The article has been sent to the editors 16.11.22
After processing 10.12.22