About Complex Objects Defining Via Integration of Data from Various Sources

Anatoly D. Khomonenko
Emperor Alexander I St. Petersburg State Transport University, 190031,
Military Space Academy name's A.F.Mozhaisky,
khomon@mail.ru

Alena I. Zimovets
Military Space Academy name's A.F.Mozhaisky,
zim_alenka@rambler.ru

Sergey P. Plyaskin
Military Space Academy name's A.F.Mozhaisky,
plyaskin.work@yandex.ru

Abstract

The analysis of up-to-day problems in information technologies is carried out. These problems are connected with increasing volumes of data in the whole world. There is considered the problem of complex object defining via integration of data from various sources. The first approach to solve this problem is to use neural-fuzzy nets. The other approach is to use invariants of the moments.

1 Introduction

We live in the age of information. Revolution progress in electronics and computer technologies leads to state when the one of tendencies of contemporary science development is increasing volumes of experimental data.[1]

The tendency of experimental plants of new generation design appears in worldwide science last days. Registration of metrological data, tasks of biological mathematics, astrological observations, aviation, power engineering and instrument making are not the only areas to discover and these branches give us colossal volumes of information. The number of various articles, including scientific ones, is growing like an avalanche. Firstly, we talk about digital publications. In other words, information published on the web pages of the authors. At the same time there are arrays of official scientific documentation including reports about work done, projects results, personal stuff information about their achievements and etc. The problems of store and processing information appear in our nowadays live. There are a lot of various information about different objects. Of course that fact is positive but at the same time it gives us problem of processing and analysis.

Analysts of IBM Company estimate Worldwide volume of data this way [Bar09]:

2003 – 5 exabytes (1 Eb = 1 billion Gb);
2008 – 0,18 zettabytes (1 Zb = 1024 Eb);
2015 – more than 6,5 zettabytes;
2020 – 40-44 zettabytes (expected);
2025 – volumes will increase in 10 times.

It is important to know how to deal with this data and know what conclusions we can make.

Today we have an opportunity to get information about one complex object from various sources. There is a question about using this data the most effectively. In this paper, we talk about how to recognize complex objects with help of information gotten from various sources.

In itself, digital format consists of concealed data that is why it is necessary to use special analysis methods. Human recognizing and analysis of information can take a lot of time. If we have giant volumes of data, the operativeness is impossible. It is clear that using of automation tools increase the effectiveness of this process. There are 2 contemporary and effective ways of complex objects identification. The one is identification via integration of data with help of neural-fuzzy nets [Kho18]. The other one is identification of complex objects via invariants [Kho16].

In: A. Khomonenko, B. Sokolov, K. Ivanova (eds.): Selected Papers of the Models and Methods of Information Systems Research Workshop, St. Petersburg, Russia, 4-5 Dec. 2019, published at http://ceur-ws.org.

Perspective calculating complexes intended to solve tasks of multichannel processing of information in real time for functional and automatic control tools.

Technical calculation complexes ensure solving following tasks [Laz16]:

- automatic processing of information from receiving tools;
- automated control of hardware complex functionality;
- representation and documentation of processing results;
- control, diagnostic, optimization of the work and ensuring stability of instruments functions.

The main stages of recognizing are collecting and processing of complex objects parameters gotten from various sources, conducting of complex objects catalogue, measures identification with catalyzed information, detection of new complex objects and definition their parameters, recognizing complex objects.

Both following methods have mathematical apparatus of fuzzy sets theory, nets and measuring and theory of planning. We can improve effectiveness and quality of recognizing and analysis via using this methods and automatic systems. We also can create complex of analytical information and knowledge with help of intelligent information technologies. This complex will give us an opportunity to get conclusions about complex objects in developing environment.

To achieve these goals, the following tasks must be solved:

1) the creation of a single functional space and indicators characterizing the state of complex objects based on a central database of information and knowledge with accumulation, storage, access and management;
2) integration existing local data bases in central information store;
3) collecting, accumulation and using of experts' knowledge in distributed bases for making conclusions and recommendations;
4) continuous observation (complex analysis) of current situation;
5) increasing effectiveness and quality of analytical instruments;
6) automation of analytical reports;
7) visualization of information via pictorial graphics;
8) expert instrumental and informational support of analytical activities.

2 Identification of complex objects via data integration with help of neural-fuzzy net

From system analysis point of view perspective system of data processing must provide complex objects classification, identification them by known signs and forecasting of complex objects development.

We can increase quality of classification and identification of complex objects with help of program means and integration of data gotten from different sources. The other method is using algorithms of fuzzy output.

We are able to get information recurred for classification from one source in insufficient volume. In conditions of incompleteness and inaccuracy of information, the construction of a mathematical model is problematic. The solution is to use few sources. It gives an opportunity to recover information. In the result, quantity of data increases.

The information integration system frees users from the need to know the data from which sources they use, what are the properties of these sources and how to access them. Sources of data can have different properties. We can choose method of integration by source properties. With an increase in the volume of information, its duplication may take place. In connection with this, with an increase in quantitative indicators, qualitative indicators may deteriorate. Approximation is able to solve this problem. The universal and commonly used method is least squares approximation.

Fuzzy output algorithms can used to recognize complex objects, actually Neural-fuzzy net. Neural-fuzzy net is a neural-fuzzy with fuzzy signals weights and activation function but at the same time with unification of x_i and w_i, p_1 and p_2 with using t-norma, t-conorma or other continuous operations [Kho18]. Inputs, outputs and weights of net are real numbers within limits of [0,1] segment. In other words, neural-fuzzy net is a neural net designed on the basis of layered architecture with using of "&", "OR" neurons.

Process of work with data based on neural net consists of three steps. There are preparation of data, extraction of the rules and estimation of them. Preparation process must define and process gotten data to make it good for specific methods of intellectual analysis. The main methods of rules

extraction are LTE (Limited Relative Error) method, black box method, method of fuzzy rules extraction from fuzzy nets, algorithm of particular rules extraction (Partial-RE), algorithm of full rules extraction (full-RE). Rules of estimation correspond following tasks. There are finding optimal sequences of rules extraction, checking of precision rules extraction, definition of knowledge quantity in neural net.

It is necessary to teach neural net before using. In deep learning we use experts' knowledge. Fecundity of identification and classification of complex objects increases when the machine learn how to do conclusions on incomplete or inaccurate information.

Learning of neural net as known is accomplished on the basis of exactly known parameters and proper output characteristics [Kho18]. If availing data for learning is not enough, information fills up via data integration from the other sources or with help of approximation. In the result, we get common functional dependence projected output characteristics of complex objects from their parameters.

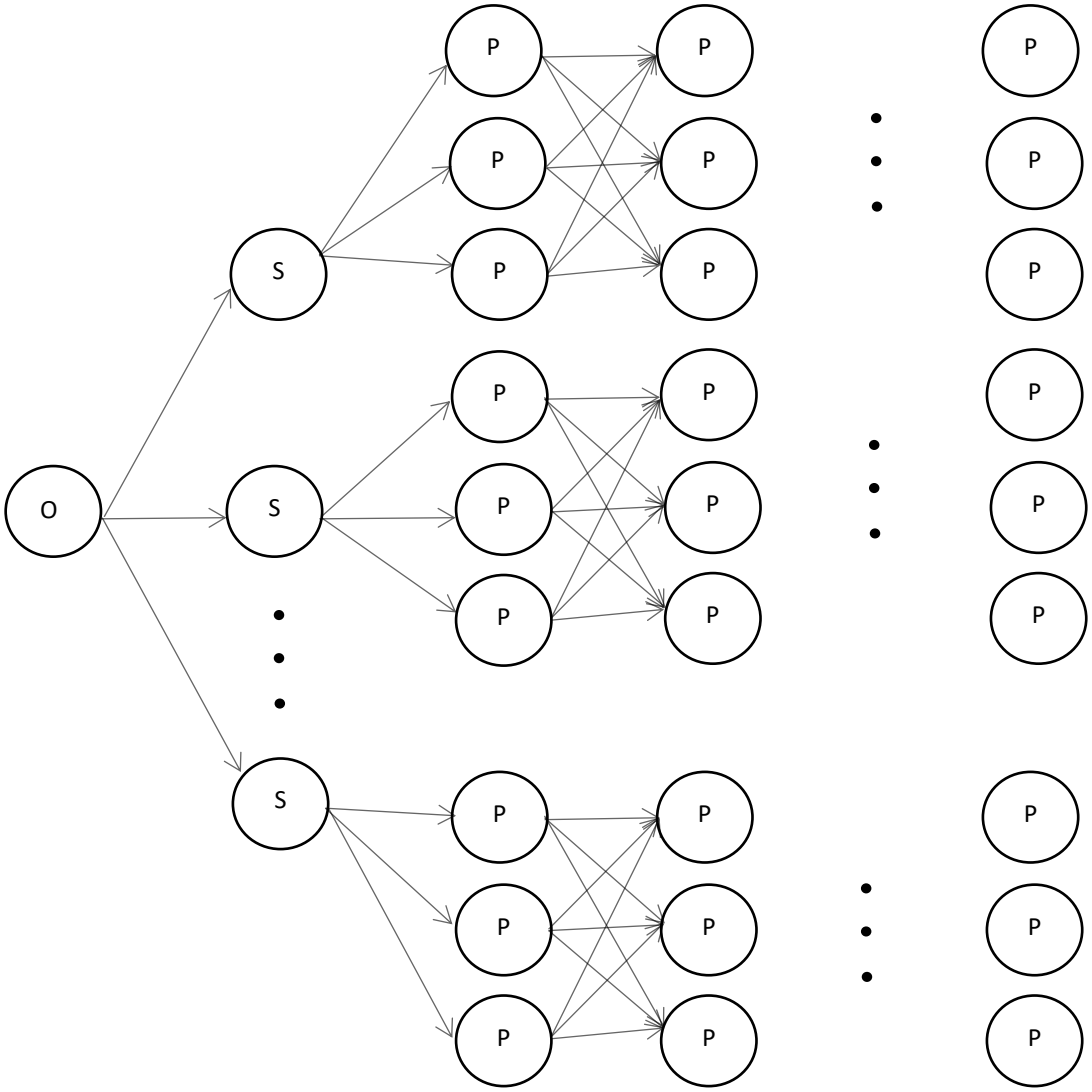

Figure 1: Net Structure (o – complex object; s – source on information; p – parameters)

By this approach, neural net turns out functional equivalent of some model of dependence between

variable like those that built on traditional modeling. The dignity of this approach is commonality and universality of admissible input parameters.

Formal task of neural net learning can be described in the following way [Kho18].

Teaching base: X = <m, n, k> – set of parameters; M = {mi | i = 1…m} – set of classes.

Variety of recognition parameters:

N = {n_j | j = 1 … n};

To find

$$K_i = \arg\max_{K_i \in K}[\hat{F}(K_i)],$$

In this case, the model of recognizing can be presented like graph (Figure 1). The graph shows the relationship between the recognition parameters of complex objects. Vertices of the graph must contain weights defining probability of complex objects recognizing with present parameters gotten with help of statistical method. A back propagation algorithm used to select the training parameters of the presented multilayer neural network.

Fuzzy output system advantage is transparency of fuzzy nets that possible by its linguistic interpretation in fuzzy production rules. Disadvantage is a priori components definition of these models [Dol14]. Advantage of neural nets is opportunity to reveal of data sequences. In other words, there is extraction of gotten knowledge. Disadvantage is complexity of net size and structure definition. Disadvantages compensated by combined using with their advantages. Linguistic structure of rules base promotes to comprehend and analyze the system.

This way neural-fuzzy net is a processor with massive parallelization of operations. This processor has essential capability to save experimental knowledge and make it accessible for following usage. Nets are similar to human brain in two ways. Net gets knowledge in learning process and use quantity of intensity inter-neuron connects.

3 Complex objects identification via invariants of the moments

Fecundity of identification and classification can be increased with using program means and integration of data gotten from various sources and with help of invariants of moments.

It is known that the set of complex objects states can be presented not only in the view of diagnostic features but with set of invariants values [Kho16].

In our case, the identification and classification tasks must considered like common theory of image recognizing of current states of the objects. Solving these tasks in technical diagnostics depends on models. Those models connect complex objects states and its representation in set of diagnostic signs. Parameters compared with calculated values estimated for every kind of complex objects to define current state in the set of parameters. Set of diagnostic signs can be considered as array X([n, p]) consisted from n measuring by p telemetric parameters.

Set of states in which complex objects are able to stay presented with help of two-dimensional invariants of the moments.

$\mathcal{S}=\{S_i \mid i=(1,m)\}$ – set of states;

$\Pi=\{\pi_j \mid j=(1,n)\}$ – set of diagnostic parameters, all states $Si \in \mathcal{S}$ pairwise distinct.

This way Si corresponded to its defined vector πj characterized by values of diagnostic signs.

In any moment of the time t in T vector πj can be presented as two-dimensional function f(x, y) where x defines time t, y is diagnostic sign, amplitude f is value of current diagnostic sign at the moment t.

This approach can be used to recognize information presented as images [Kho16]. Relevance of this problem consists in the fact that the main part of information processed by human in process of discovering design and control activity is graphic information like photographs, diagrams, sketches and drafts. Apparatus of imagine representation and description based on invariants calculation is needed to solve identification tasks.

Central moment of digital image f(x,y) described

$$\mu_{pq} = \sum_x \sum_y (x-\bar{x})^p (y-\bar{y})^q f(x,y),$$

were $\bar{x} = m_{10,} / m_{00}, \bar{y} = m_{01} / m_{00}.$

In the process of analysis seven invariants are used. The parts of considered image are invariant to transfer turning axial symmetry pressing and stretching.

$$\varphi_1 = \eta_{20} + \eta_{02};$$

$$\varphi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2;$$

$$\varphi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2;$$

$$\varphi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})((\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2) +$$
$$+(3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})(3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2);$$

$$\varphi_6 = (\eta_{20} - \eta_{02})((\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2) +$$
$$+4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03});$$

$$\varphi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})((\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2) -$$
$$-(\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{30})(3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2).$$

In itself image recognizing consists of two relatively independent tasks. The first is classification of groups based on the adjusted requests. The second is taking object to the one of them.

When solving the first task it needed to prepare data and form set of states for diagnostic signs. It is necessary that forms of representation diagnostic data and Si and πj vectors to be correspondent. It can be discrete or analog forms.

When solving the second task it is required comparison between reference values and calculated invariants. If these values do not diverge from reference ones more than allowed quantity, then the state is defined.

Thus we can conclude that complex objects defining via invariants of the moments is a compound task. We can also say that this method is very effective in analysis of information gotten from sources in image way.

4 An example of a neural-fuzzy network for classifying objects

Let's consider an example of constructing a neural-fuzzy network ANFIS **[Jyh93]** in MATLAB to classify space debris objects. In table 1 a fragment of the training set for the neural-fuzzy network ANFIS presented.

Table 1: A fragment of the training Set

T	i	Ha	Hp	Exit
92.5935	51.65	422	403	1
91.22653	51.56	324	333	1
170.9687	53.32	5185	4367	2
98.4683	43.56	301	344	2
92.5944	51.66	416	456	2
189.367	63.34	1991	1286	2
104.046	98.15	712	69S	3
96.1496	94.11	38654	35120	3
98.9342	65.02	981	678	3
93.7549	74.75	354	454	3
104.975	74.03	1016	243	3
718.157	69.68	11234	6918	3
1437.62	13.77	594	578	4

Input Variables:

- **T** is the longitude of the ascending angle;
- **i** is the inclination;
- **Ha** – semi major axis;;
- **Hp** – pericenter.

The output variable corresponds to various elements of space debris:

1– Idle satellites.
2 – Overclocking blocks.
3 – Parts of space rockets.
4 – Worked out steps.
5 – Technological elements.
6 – Wreckage.
7 – Small element (up to 10 cm).

The ANFIS Editor with training set shown on the fig. 2.

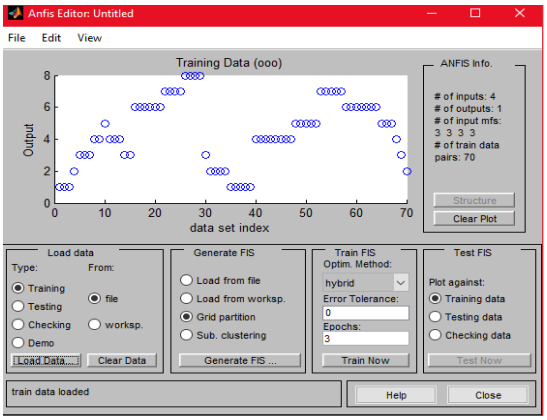

Figure 2: ANFIS Editor with training set

The graph of the dependence of learning errors on the number of training cycles presented on the fi. 3.

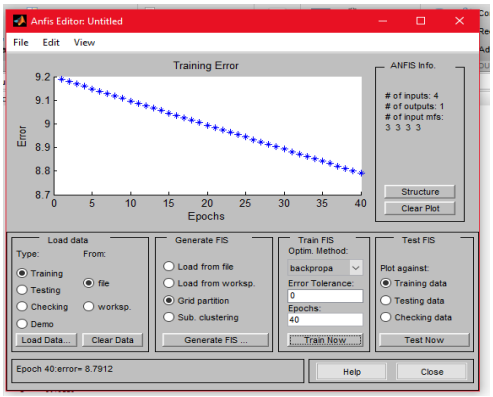

Figure 3: The dependence of learning errors on the number of training cycles

The given example shows the possibility of using the ANFIS fuzzy neural network to solve the problem of space debris classification when integrating data from various sources.

5 Conclusion

It is considered two ways of solving problems of complex objects recognizing by integration of data from different sources. The one of them is using neural-fuzzy nets. That method gives us an opportunity to analyses data consisted of incomplete information. Another method is considered as using invariants of the moments. This method is the most effective recognizing complex objects in image form.

Informational technologies influence on various scientific fields. Nowadays instrumental tools develop in very active way. This tools give an opportunity to get information about different complex objects like human genome or data about distant star systems. We can get detailed definition about all possible objects and concepts by big data analysis.

Proposed directions of development give us substantial building up possibilities of information recognizing systems. Firstly it touches on efficiency of complex objects identification and classification.

Acknowledgments

The work was partially supported by the grant of the MES RK: project No. AP05133699 "Research and development of innovative information and telecommunication technologies using modern cyber technical means for the city's intelligent transport system".

References

[Kho18] A D. Khomonenko, E. L. Yakovlev (2018) The rationale for the architecture of the convolutional neural network for object recognition on images on-board computer system. H&ES Research. №Vol. 10. No. 6, Pp. 86-93.

[Kho16] A D. Khomonenko, E. L. Yakovlev (2016) State estimation of complex object via invariants. 3RD Russian conference "Mathematical modeling and information technologies", MMIT, Ceur Workshop Proceeding. Pp.44-49.

[Dmi87] A.K. Dmitryev, R.M. Yusupov (1987) Identification And Technical Diagnostics. Textbook for universities. Leningrad.: EM USSR.521 p.

[Laz16] Lazutin O.G. (2016) The technique of bringing information about the technical condition of space vehicles using data compression algorithms and cognitive graphical representation. – SPb.: MSA name's A.F.Mozhaisky, p. 11-17.

[Kop15] E.V. Kopkin, V.A. Chikurov, V.V. Aleynik, O.G. Lazutin (2015) Algorithm for constructing a flexible program for technical object diagnosing on the criterion of received information value. SPIIRAS Proceedings. Vol. 4(41). – Pp. 106-130.

[Kop15] Kopkin E.V., Kravcov A.N., Lazutin O.G. (2015) Selection of discrete diagnostic features taking into account their value for the recognition of the technical condition of the object. Information And Space. № 2. Pp. 111-117.

[Bub15] V.P. Bubnov, M.L. Gluharev, A.A. Kornirnko, S.A. Krasnov, V.V. Rogalchuk, A.V. Tyrva, V.V. Fedyanin, A.D. Khomonenko (2015) Models of information systems. Textbook. Moscow, 2015.

[Isa13] E.A. Isaev, V.V. Kornilov (2013) The Problem of Processing and Storage of Large Amounts of Scientific Data and Approaches to Its Solution. Mathematical Biology and Bioinformatics. Vol. 8, №1, Pp.49-65

[Man11] V.G. Manzhula, D.S. Fedjashov (2011) Kohonen Neural Networks And Fuzzy Neural Networks in Data Mining. Fundamental Research. № 4. p. 108-115

[Yan16] M.V. Yanaeva, E.V. Sinchenko (2016). Indistinct neural nets in data mining. Scientific works of KubSTU. № 15. Pp.80-93

[Kog10] M.R. Kogalovsky (2010). Methods of Data Integration in Information Systems. Electronic Publications of the RAS p.9

[Dol14] O.N. Dolinina, A.Y.Shvarts (2014). Several aspects of increasing of the effectiveness of the making decisions by intelligent systems using fuzzy inference. Fundamental research. №12. Pp. 691-695.

[Bar09] A.A. Bargesjan (2009). Data and process analysis. Textbook. BHV-Petersburg. 512 p.

[Jyh93] Jyh-Shing, Roger Jang. ANFIS: adaptive-network-based fuzzy inference system. IEEE Transactions on Systems, Man and Cybernetics : journal. 1993. Vol. 23, No. 3. DOI:10.1109/21.256541.