Bibliography

1. Koval V.N., Savyak V.V., Sergienko I.V., "Tendencies of modern supercomputer systems development", Control Systems and Computers, Vol.6, November-December 2004, 31-44 pp (In Russian).
2. Koval V.N., Savyak V.V., "Multiprocessor cluster systems: planning and implementation", "Nauka osvita", Artificial Intellect, Vol.3, 2004, 117-126 pp (In Russian).
3. www.gromacs.org
4. Koval V., Bulavenko O, Rabinovich Z Parallel Architectures and Their Development on the Basis of Intelligent Solving Machines // Proc. Int. Conf. on Parallel Computing in Electrical Engineering. — Warsaw (Poland).- 2002.- P.21-26

Authors' Information

Valeriy N. Koval – Institute of Cybernetics NAS Ukraine; Prospekt Academika Glushkova,40, Kiev, 03680 MCP, Ukraine; e-mail: icdepval@ln.ua

Sergey G. Ryabchun – Institute of Cybernetics NAS Ukraine; Prospekt Academika Glushkova,40, Kiev, 03680 MCP, Ukraine; e-mail: sr@emt.com.ua

Volodymyr V. Savyak – Institute of Cybernetics NAS Ukraine; Prospekt Academika Glushkova,40, Kiev, 03680 MCP, Ukraine; e-mail: Volodymyr.Savyak@ustar.kiev.ua

Ivan V. Sergienko – Institute of Cybernetics NAS Ukraine; Prospekt Academika Glushkova,40, Kiev, 03680 MCP, Ukraine

Anatoliy A. Yakuba – Institute of Cybernetics NAS Ukraine; Prospekt Academika Glushkova, 40, Kiev, 03680 MCP, Ukraine; e-mail: ayacuba@voliacable.com

NEW KNOWLEDGE OBTAINING IN STRUCTURAL-PREDICATE MODELS OF KNOWLEDGE

Valeriy Koval, Yuriy Kuk

Abstract: An effective mathematical method of new knowledge obtaining on the structure of complex objects with required properties is developed. The method comprehensively takes into account information on the properties and relations of primary objects, composing the complex objects. It is based on measurement of distances between the predicate groups with some interpretation of them. The optimal measure for measurement of these distances with the maximal discernibleness of different groups of predicates is constructed. The method is tested on solution of the problem of obtaining of new compound with electro-optical properties.

Keywords: New knowledge, Predicates, Complex objects, Primary objects, Maximal discernibleness.

ACM Classification Keywords: I.2.4 Artificial Intelligence: knowledge representation formalisms and methods.

Introduction

The present work deals with further development of methods of practical extraction of knowledge from experimental data. Its purpose is development of an effective mathematical method for obtaining of new knowledge on the structure of complex objects with certain properties. The work is focused on solution of an important applied problem - designing of structure of compounds with the needed properties.

In our previous works [1] - [2] in order to obtain new knowledge in form of production rules, a concept of variable predicate, able to accept a number of values - so-called predicate constants, predicates in the conventional sense

- as well as a concept of distance between the predicates were used. Both these concepts were further developed in the present work. However unlike the above mentioned works the present article considers predicates with the subject domains consisting of the objects, possessing an internal structure are considered, that is objects from subject domains of the predicates are considered as complex while earlier they were considered as integral. We will refer to components of complex object as primary objects [3], and to the predicates designating properties and the relations of primary objects as primary predicates. Normally, some information is also known on properties and relations of primary objects being parts of complex objects, which should be used in the procedures of new knowledge obtaining on the structure of complex objects possessing certain properties. The procedure of such knowledge obtaining proposed in the work is based on measurement of the distances between the groups of properties and relations between the primary objects, or in terms of logic, between the groups of predicates with some interpretation of them. The measure introduced by us in the works [1] - [2] for measurement of the predicate affinity degree, cannot be directly transferred on the groups of predicates. Therefore, in the present work a special measure, an optimal one in terms of maximal discernibleness of different groups of predicates for measurement of distances between them is introduced.

1. Structural-predicate Model of Knowledge

It is convenient to represent knowledge of complex objects in the form which we have named *the structural - predicate model of knowledge*. It is a further generalization of the structural - attributive model of knowledge [3]-[4]. The generalization is in that their relations, rather than just the properties of the objects are also considered. For example, the two-place predicate «difference of melting temperatures of two substances is more than Δ» describes a certain relation between two objects.

The structural - predicate model of knowledge (SPMK) is a four-layer graph of a pyramidal network, separate layers of which form its nodes. For clearness fig. 1 represents SPMK, containing knowledge on properties of chemical compounds with various types of crystal lattice structure: such as LiCaAlF6 (L-structure of the lattice), Na2SiF6 (N-structure of the lattice), Trirutile (T-structure of the lattice). Let us designate as P, A, S, V following sets of SPMK nodes. The first layer P corresponds to the predicate constants (values of variable predicates), designating properties and relations of the primary objects. Let the elements P be *primary predicates*. On fig. 1 primary variable predicates are: Tm - melting point, So - standard entropy for corresponding simple oxides, H - standard enthalpy formations for corresponding simple oxides, Rs - radius of ions, C - isobaric thermal capacity. On fig. 1 each of these predicates takes 2 values, and predicate constants corresponding to them are designated with figures 1 and 2.

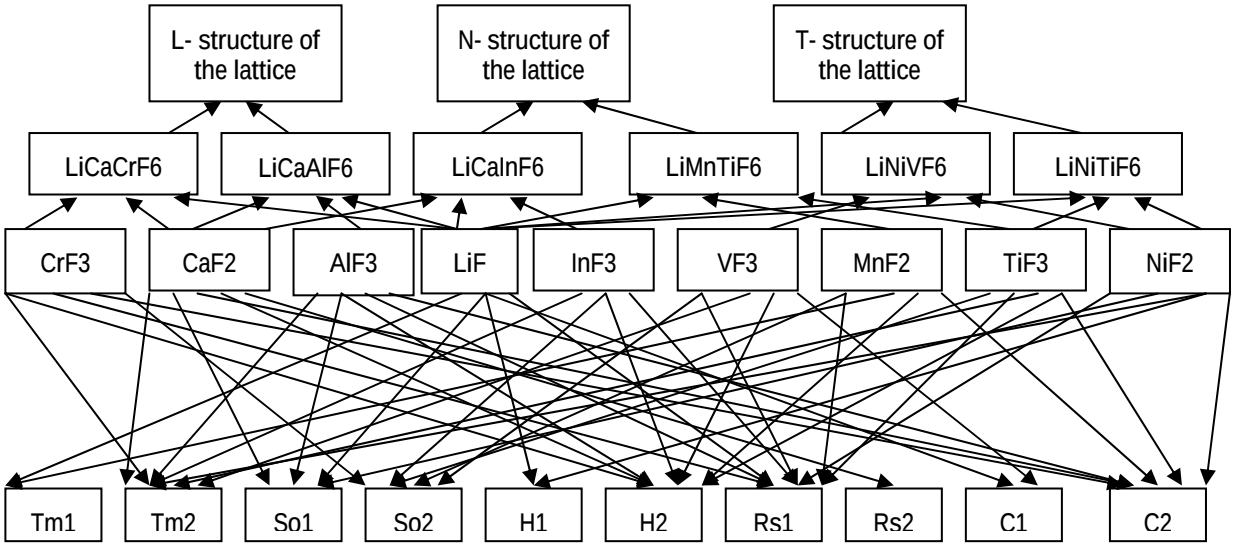

Fig. 1 Example of the structural - predicate model of knowledge

The second layer A corresponds to the names of the primary objects. They compose subject domains of primary predicates at their interpretation. The third layer S corresponds to the names of the compound objects, the fourth layer V - to the values of the variable predicates designating properties and relations of the compound objects. Let us call elements V *predicates of compound objects*. Their subject domains are compound objects. On fig. 1 predicates of compound objects are 3 values of the variable predicate «to have a certain type of the crystal lattice». Arches of the bottom and top layers connect the nodes representing objects to the nodes representing predicate constants and are directed from the primary and compound objects to the predicate constants. They are used in interpretation of the predicates. Let ω designate multiplicity of some predicate constant. Then the presence of the ω arches originating from ω objects and converging in a certain predicate corresponds to the logical value of the predicate "true" in substitution of these objects into the predicate, and to the value "false" in substitution of the object into the predicate in case of absence of the arch, connecting the object with the predicate. Thus, in substitution of the objects A and S from which the arrows go to the predicate constants of the sets P and V, two sets of knowledge R_1 and R_2 in form of true statements on properties and relations of the primary and compound objects are formed instead of the arguments of these predicate constants.

The arches of the middle layer connect the nodes corresponding to the primary objects to the nodes representing the compounds. The primary elements, from which the arches originate, are parts of those complex objects in which these arches terminate.

2. Measure of Affinity of the Predicate Groups

Let's consider a problem of construction of a measure for measurement of degree of affinity of variable predicate groups. This measure should possess the following natural property: the distance between the groups of predicates, measured with it, will not be zero, when these groups of predicates are different. It follows, that it will possess a property of maximal discernibleness of different groups of predicates. Let us construct a measure satisfying this property.

Let's designate N - number of primary predicates in structural - predicate model, M - number of predicates of compound objects, $n(k)$ - number of the primary objects, which are included in complex object s_k. With symbols $p_1,\cdots p_N$ we designate primary variable predicates of the model. In case of numerical values of variable predicates we will base on the following rule: indexes for their predicate constants are selected in the way that their sequence order corresponds to the sequence order of numerical values of the variable predicates. Thus when dividing the interval of variation of numerical values of predicates into segments (quantization), the predicate constant indexes, corresponding to them, will coincide with numbers of these segments.

Definition 1. *Let us understand as the label* x_{ik} *of a primary variable predicate* p_i *for complex object* s_k an index of the predicate constant of the predicate p_i which takes the logical value "True" when the primary objects included into s_k and connected by arches with this predicate constant are substituted into it as arguments.

Let us call a vector, the elements of which are labels for complex object s_k of all primary predicates which are included into the structural - predicate model of knowledge, a *distribution of labels* $x_k = (x_{1k}, x_{2k}, \cdots, x_{Nk})$ *of primary predicates for compound* s_k. Let us name the vector $h^{(1)} = (\bar{x}_1^{(1)}, \bar{x}_2^{(1)}, \cdots, \bar{x}_N^{(1)})$, the coordinates of which are equal to average values of components of vectors of labels in this group, a *typical label distribution* for the complex object group $G_1 = \left\{s_1^{(1)}, s_2^{(1)}, \cdots, s_K^{(1)}\right\}$. Let us name the vector $\tilde{x}_k = (x_{1k} - \bar{x}_1^1, x_{2k} - \bar{x}_2^1, \cdots, x_{Nk} - \bar{x}_N^1)$ *a centralized vector of labels of primary predicates for the complex object* s_k belonging to the group of complex objects $G_1 = \left\{s_1^{(1)}, s_2^{(1)}, \cdots, s_K^{(1)}\right\}$.

Primary predicate label distributions for a certain group of complex objects represent a set of points in the space R_N. If there are two groups of complex objects, we will obtain two such sets of points, which are intermixed in a random way. Following problem emerges: it is required to find the characteristics of these sets,

which will allow measuring the degree of affinity of the groups with them. A following solution suggests itself: to take as characteristics of each set of points corresponding points with the average coordinates which represent the typical distribution of the labels for corresponding groups, and to take as a measure of affinity for these groups an Euclidean length of the distance between them. However, this decision is not the optimal one. Indeed, let us consider a following example. Let points of distributions of the labels of the primary predicates for both groups be located on two perpendicular straight lines symmetrically to the their intersection point, with the points of every group lying on the separate line. It can be easily seen, that typical distributions of labels for both groups will coincide, and, hence, the distance between them is zero although the groups of predicates themselves are different. The non-optimal solution above can be seen from another point, allowing to find the optimal one. Let us draw a direct line through two points which are in R_N typical distributions of labels of the primary predicates for complex objects of both groups, and project on it the sets of points for both groups. It is easy to see that average values of projections for both sets of points, the so-called centers of projections $\bar{z}^1$ and $\bar{z}^2$ of corresponding groups, coincide with the points representing typical distributions of labels. As the measure should possess the property of maximal discernibleness of different groups of predicates a following optimization problem arises: to construct in R_N the straight line c, not necessarily going through typical distributions labels, such that the distance between the centers of projections of both sets of points is the maximal one. The criterion of optimization of this problem is: $\bar{z}^1 - \bar{z}^2 \to \max$. The distance $\bar{z}^1 - \bar{z}^2$ obtained in result of optimization should be taken as a measure of affinity of complex objects for both groups.

3. Mathematical Apparatus for Construction of the Measure

It follows from the previous section, that in order to find an optimal solution of the problem of construction of the measure for degree of affinity of groups of variable predicates it is necessary to construct some auxiliary straight line c in space R_N and to project distributions of the labels for both groups of predicates on it. As a result, we will receive two overlapping sets of points on the straight line. As the choice of direction of the line c influences the distances between the projections of distributions of the labels and, hence, on their affinity the straight line should be chosen in the way that the projections of distributions of the labels from different groups of the complex objects are removed from each other as far as possible. Such choice of direction of the straight line will allow distinguishing different groups of complex objects in the optimum way.

Let us name the straight line c, on which distributions of the labels of the primary predicates for complex objects are projected *a projective* line.

Let's cite without proofs a number of auxiliary statements needed to find the required projective line.

Lemma 1. Projection of distributions of the labels $x_k = (x_{1k}, x_{2k}, \cdots, x_{Nk})$ of primary predicates for complex object s_k on the projective line c, going through the point of origin in the space R_N, is defined by the formula $\Pr_c x_1 = c_1 x_{11} + c_2 x_{21} + \cdots + c_N x_{N1}$, where $(c_1, c_2, \cdots, c_N)$ are cosines of the angles, formed by a straight line with the coordinate axis.

Definition 2. *Let us call the total distance* $D_1(z)$ concerning any point z of projections of distributions of labels of primary predicates for group of complex objects $G_1 = \left\{s_1^{(1)}, s_2^{(1)}, \cdots, s_K^{(1)}\right\}$.

$$D_1(z) = \sum_{\nu=1}^{K} \left\| z_\nu^{(1)} - z \right\|, \text{ where } z_1^{(1)} = c_1 x_{11}^{(1)} + \cdots + c_N x_{N1}^{(1)}, \ldots, z_K^{(1)} = c_1 x_{1K}^{(1)} + \cdots + c_N x_{NK}^{(1)}.$$

A scattering in relation to an arbitrary point z of projections of distributions of the labels of primary predicates for group of complex objects $G_1 = \left\{s_1^{(1)}, s_2^{(1)}, \cdots, s_K^{(1)}\right\}$

Let us call an average value of projections of distributions of labels of primary predicates for the group G_1: $\bar{z}^{(1)} = \frac{1}{K}\sum_{\nu=1}^{K} z_\nu^{(1)}$ a *center of projections* for the group of complex objects $G_1 = \left\{s_1^{(1)}, s_2^{(1)}, \cdots, s_K^{(1)}\right\}$

Lemma 2. *Scattering in relation of an arbitrary* point z of projections of distributions of the labels of primary predicates for group of complex objects $G_1 = \left\{s_1^{(1)}, s_2^{(1)}, \cdots, s_K^{(1)}\right\}$ is minimal, when z is equal to the center of their projections: $z = \bar{z}^{(1)}$, thus $D_1(z) = D_1(\bar{z}) = \sum_{\nu=1}^{K}(z_\nu^{(1)} - \bar{z}^{(1)})^2$.

Let $G_1 = \left\{s_1^{(1)}, s_2^{(1)}, \cdots, s_K^{(1)}\right\}$ and $G_2 = \left\{s_1^{(2)}, s_2^{(2)}, \cdots, s_L^{(2)}\right\}$ - two groups of the complex objects consisting correspondingly from K and L complex objects. For each complex object of these groups, on the basis of the structural - predicate model of knowledge we will construct distribution of labels of its primary predicates. We will obtain $K + L$ vectors, which in the space R_N will be represented with two sets of the vectors: X_1 - the set of vectors $x_1^{(1)}\ x_2^{(1)}$, ..., $x_K^{(1)}$ and X_2 - the set of vectors $x_1^{(2)}\ x_2^{(2)}$, ... $x_L^{(2)}$.

Let us project sets X_1 and X_2 on a projective line c. On the basis of lemma 1, we will find the values of projections:

$z_1^{(1)} = \Pr_c x_1^{(1)} = c_1 x_{11}^{(1)} + c_2 x_{21}^{(1)} + \cdots + c_N x_{N1}^{(1)}$, ..., $z_1^{(2)} = \Pr_c x_1^{(2)} = c_1 x_{11}^{(2)} + c_2 x_{21}^{(2)} + \cdots + c_N x_{N1}^{(2)}$,

$z_2^{(2)} = \Pr_c x_2^{(2)} = c_1 x_{12}^{(2)} + c_2 x_{22}^{(2)} + \cdots + c_N x_{N2}^{(2)}$, ... $z_L^{(2)} = \Pr_c x_L^{(2)} = c_1 x_{1L}^{(2)} + c_2 x_{2L}^{(2)} + \cdots + c_N x_{NL}^{(2)}$.

Let's designate sets of projections X_1 and X_2 as Z_1 and Z_2 accordingly, and their centers:

$$\bar{z}^{(1)} = \frac{1}{K}(z_1^{(1)} + z_2^{(1)} + \cdots + z_K^{(1)}),\ \bar{z}^{(2)} = \frac{1}{L}(z_1^{(2)} + z_2^{(2)} + \cdots + z_L^{(2)}).$$

Definition 3. Let us call scattering in relation to an arbitrary point z of projections of distributions of labels of primary predicates for the combined group of the complex objects $G = G_1 \bigcup G_2$ *a total scattering* of both groups $D(z)$.

It is obvious, that it is equal $D(z) = \sum_{\nu=1}^{K}\left\|z_\nu^{(1)} - z\right\| + \sum_{\nu=1}^{L}\left\|z_\nu^{(2)} - z\right\|$.

Let call *a common center* of the combined set of projections $Z = Z_1 \bigcup Z_2$ the value $\bar{z} = \frac{1}{K+L}(z_1^{(1)} + \cdots + z_K^{(1)} + z_1^{(2)} + \cdots + z_L^{(2)})$.

Let's cite without proof following theorem on scattering of projections of distributions of primary predicate labels.

Theorem 1. Total scattering $\overline{D} = \overline{D}(\bar{z})$ in relation to the common center $\bar{z}$ of projections of distributions of the labels of primary predicates of the combined group of complex objects $G = G_1 \bigcup G_2$ is calculated according to the formula $\overline{D} = \overline{D}_1 + \overline{D}_2 + \hat{D}_1 + \hat{D}_2$, where

$$\overline{D}_1 = \sum_{\nu=1}^{K}(z_\nu^{(1)} - \bar{z}^{(1)})^2,\ \overline{D}_2 = \sum_{\nu=1}^{L}(z_\nu^{(2)} - \bar{z}^{(2)})^2,\ \hat{D}_1 = K(\bar{z}^1 - \bar{z})^2,\ \hat{D}_2 = L(\bar{z}^{(2)} - \bar{z})^2.$$

It follows from theorem 1 that in order to maximize the measure of discernibleness of both groups of predicates – the distance $\bar{z}^1 - \bar{z}^2$ - it is necessary to maximize the sum $\hat{D}_1 = K(\bar{z}^1 - \bar{z})^2$ and $\hat{D}_2 = L(\bar{z}^2 - \bar{z})^2$. Let us call the sum $D_1 + D_2$ *a complete discernibleness* .

Let us obtain expressions for complete discernibleness and the sums of scattering of label distribution projections of both groups of predicates, which are used for the further calculations.

The vector of the difference of typical distributions of labels for the groups of complex objects G_1 and G_2 let us designate as $h = h^{(1)} - h^{(2)} == (\bar{x}_1^{(1)} - \bar{x}_1^{(2)}, \bar{x}_2^{(1)} - \bar{x}_2^{(2)}, \cdots, \bar{x}_N^{(1)} - \bar{x}_N^{(2)})$.

Let's construct a square matrix $H = h^T h$ where the top index T designates operation of transposing. Dimension H is $N \times N$. It looks like

$$H = \begin{pmatrix} (\bar{x}_1^{(1)} - \bar{x}_1^{(2)})^2 & (\bar{x}_1^{(1)} - \bar{x}_1^{(2)})(\bar{x}_2^{(1)} - \bar{x}_2^{(2)}) & \cdots & (\bar{x}_1^{(1)} - \bar{x}_1^{(2)})(\bar{x}_N^{(1)} - \bar{x}_N^{(2)}) \\ (\bar{x}_2^{(1)} - \bar{x}_2^{(2)})(\bar{x}_1^{(1)} - \bar{x}_1^{(2)}) & (\bar{x}_2^{(1)} - \bar{x}_2^{(2)})^2 & \cdots & (\bar{x}_2^{(1)} - \bar{x}_2^{(2)})(\bar{x}_N^{(1)} - \bar{x}_N^{(2)}) \\ \cdots & \cdots & \cdots & \cdots \\ (\bar{x}_N^{(1)} - \bar{x}_N^{(2)})(\bar{x}_1^{(1)} - \bar{x}_1^{(2)}) & (\bar{x}_N^{(1)} - \bar{x}_N^{(2)})(\bar{x}_2^{(1)} - \bar{x}_2^{(2)}) & \cdots & (\bar{x}_N^{(1)} - \bar{x}_N^{(2)})^2 \end{pmatrix}.$$

Let's consider a matrix H' with elements $h'(\nu,\mu) = \dfrac{KL}{K+L} h(\nu,\mu)$ and designate $A^{(1)}$ and $A^{(2)}$ matrixes the columns of which will consist of components of *centralized* primary predicate label distributions for corresponding groups of complex objects. They contain N lines and accordingly K and L columns.

$$A^{(1)} = \begin{pmatrix} x_{11}^{(1)} - \bar{x}_1^{(1)} & \cdots & x_{1K}^{(1)} - \bar{x}_K^{(1)} \\ \cdots & \cdots & \cdots \\ x_{N1}^{1} - \bar{x}_1^{1} & \cdots & x_{NK}^{(1)} - \bar{x}_K^{(1)} \end{pmatrix}, \; A^{(2)} = \begin{pmatrix} x_{11}^{(2)} - \bar{x}_1^{(2)} & \cdots & x_{1L}^{(2)} - \bar{x}_{lK}^{(2)} \\ \cdots & \cdots & \cdots \\ x_{N1}^{(2)} - \bar{x}_1^{(2)} & \cdots & x_{NL}^{(2)} - \bar{x}_L^{(2)} \end{pmatrix}$$

Let's consider matrixes $B^{(1)} = A^{(1)} A^{(1)^T}$ and $B^{(2)} = A^{(2)} A^{(2)^T}$. They look like:

$$B^{(1)} = \begin{pmatrix} \sum_{\nu=1}^{K} (x_{1\nu}^{(1)} - \bar{x}_1^{(1)})^2 & \sum_{\nu=1}^{K} (x_{1\nu}^{(1)} - \bar{x}_1^{(1)})(x_{2\nu}^{(1)} - \bar{x}_2^{(1)}) & \cdots & \sum_{\nu=1}^{K} (x_{1\nu}^{(1)} - \bar{x}_1^{(1)})(x_{N\nu}^{(1)} - \bar{x}_N^{(1)}) \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{\nu=1}^{K} (x_{N\nu}^{(1)} - \bar{x}_N^{(1)})(x_{1\nu}^{(1)} - \bar{x}_1^{(1)}) & \sum_{\nu=1}^{K} (x_{N\nu}^{(1)} - \bar{x}_N^{(1)})(x_{2\nu}^{(1)} - \bar{x}_2^{(1)}) & \cdots & \sum_{\nu=1}^{K} (x_{N\nu}^{(1)} - \bar{x}_N^{(1)})^2 \end{pmatrix},$$

$$B^{(2)} = \begin{pmatrix} \sum_{\nu=1}^{L} (x_{1\nu}^{(1)} - \bar{x}_1^{(1)})^2 & \sum_{\nu=1}^{L} (x_{1\nu}^{(1)} - \bar{x}_1^{(1)})(x_{2\nu}^{(1)} - \bar{x}_2^{(1)}) & \cdots & \sum_{\nu=1}^{L} (x_{1\nu}^{(1)} - \bar{x}_1^{(1)})(x_{N\nu}^{(1)} - \bar{x}_N^{(1)}) \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{\nu=1}^{L} (x_{N\nu}^{(1)} - \bar{x}_N^{(1)})(x_{1\nu}^{(1)} - \bar{x}_1^{(1)}) & \sum_{\nu=1}^{L} (x_{N\nu}^{(1)} - \bar{x}_N^{(1)})(x_{2\nu}^{(1)} - \bar{x}_2^{(1)}) & \cdots & \sum_{\nu=1}^{L} (x_{N\nu}^{(1)} - \bar{x}_N^{(1)})^2 \end{pmatrix}$$

In following theorems, formulas for calculation of the complete discernibleness and the sum $\overline{D}_1 + \overline{D}_2$ of scatterings are given. The theorems are cited without proof.

Theorem 2. The complete discernibleness is equal to:

$$\hat{D}_1 + \hat{D}_2 = \frac{KL}{K+L} \sum_{\nu=1}^{N} \sum_{\mu=1}^{N} c_\nu c_\mu h(\nu,\mu),$$

where $h(\nu,\mu)$ is an element of the matrix H situated in ν line and μ column.

Theorem 3. The sum of scattering of $\overline{D}_1 + \overline{D}_2$ primary predicate label distribution projections is $\overline{D}_1 + \overline{D}_2 = \sum_{\nu=1}^{N}\sum_{\mu=1}^{N} c_\nu c_\mu b(\nu,\mu)$, where $b(\nu,\mu)$ are elements of the matrix $B = B^{(1)} + B^{(2)}$.

We cite without proof the basic theorem, which allows distinguishing groups of predicates in the way that complete discernibleness is as great as possible.

Theorem 4. Complete discernibleness $\hat{D}_1 + \hat{D}_2$ reaches its maximum at the fixed value of the sum of scatterings of groups when values of the cosines $(c_1, c_2, \cdots, c_N)$ of the pointing angles formed by the projective line with coordinate axis are the components of an eigen vector W for a nonzero eigen value of the matrix $B^{-1}H'$.

6. Stages of Development of the Structure of Complex Objects

Let's consider stages of solution of the problem on development of the structure of the compounds with set properties [3]. For clearness we will accompany this solution with an example based on the data, cited in [4]. Let it be required to develop new compounds possessing electro-optical properties. It is known, that the fluoride crystals, which have crystal structures of types LiCaAlF6 and Na2SiF6, possess electro-optical properties, and the fluoride crystals with the structure such as Trirutile do not possess these properties. For the sake of simplicity, we will limit ourselves to consideration of compounds with structures such as Na2SiF6 and Trirutile. On the first stage of development SPMK, describing the properties of fluoride compounds, is constructed. The fragment of such model is shown in fig. 1. On the second stage we single out in the SPMK a set of predicate constants of compound objects V^+, to which the required properties of projected compound correspond, and the set of predicate constants V^-, to which undesirable properties of projected compound correspond. In our example V^+ is a predicate «to have a structure such as Na2SiF6», and V^- is a predicate «to have a structure such as Trirutile». On the third stage we single out in the SPMK the set of nodes of the group G_1 corresponding to known compounds which have connections with predicates of set V^+, and have no connections with predicates of set V^-, and set of the nodes G_2 corresponding to known compounds which have connections with predicates of set V^-, and have no connections with predicates of set V^+. Let the first group include 10 compounds which are listed in the first column of the table 1, and the second group include 17 compounds listed in the first column of table 2.

Table 1

LiMgAlF6	MgF2	AlF3	1536	13,68	268,7	0,72	14,72	1545	15,8	361	0,39	17,95
LiMnAlF6	MnF2	AlF3	1133	22,25	202,4	0,83	16,24	1545	15,8	361	0,39	17,95
LiCaInF6	CaF2	InF3	1691	16,36	291,8	1	16,02	1445	33,5	250	0,8	15,93
LiMnTiF6	MnF2	TiF3	1133	22,25	202,4	0,83	16,24	1500	21,1	342	0,67	15,93
LiMnVF6	MnF2	VF3	1133	22,25	202,4	0,83	16,24	1679	23,1	271	0,64	21,62
LiMnCrF6	MnF2	CrF3	1133	22,25	202,4	0,83	16,24	1677	22,5	277	0,61	18,82
LiMnRhlF6	MnF2	RhF3	1133	22,25	202,4	0,83	16,24	1460	26	175	0,66	15,93
LiFeGaF6	FeF2	GaF3	1375	20,79	158	0,78	16,28	1225	28	255	0,62	15,93
LiCoInF6	CoF2	InF3	1400	19,59	159,1	0,745	16,44	1445	33,5	250	0,8	15,93
LiNiInF6	NiF2	InF3	1430	17,6	157,2	0,69	15,31	1445	33,5	250	0,8	15,93
		h1	1309	19,92	204,7	0,808	15,99	1496	25,30	279,2	0,64	17,33

Table 2

LiMgCrF6	MgF2	CrF3	1536	13,68	268,7	0,72	14,72	1677	22,5	277	0,615	18,82
LiMgGaF6	MgF2	GaF3	1536	13,68	268,7	0,72	14,72	1225	28	255	0,62	15,93
LiMgRhF6	MgF2	Rh3	1536	13,68	268,7	0,72	14,72	1460	26	175	0,665	15,93
LiNiTiF6	NiF2	TiF3	1430	17,6	157,2	0,69	15,31	1500	21,1	342,2	0,67	15,93
LiNiVF6	NiF2	VF3	1430	17,6	157,2	0,69	15,31	1679	23,1	271	0,64	21,62
LiCoCrF6	CoF2	CrF3	1400	19,59	159,1	0,745	16,44	1677	22,5	277	0,615	18,82
LiCuCrF6	CuF2	CrF3	1043	16,4	128,5	0,73	16,8	1677	22,5	277	0,615	18,82
LiZnCrF6	ZnF2	CrF3	1148	17,61	183	0,74	15,69	1677	22,5	277	0,615	18,82
LiNiFeF6	NiF2	FeF3	1430	17,6	157,2	0,69	15,31	1300	25	239	0,645	15,93
LiNiCoF6	NiF2	CoF3	1430	17,6	157,2	0,69	15,31	1230	27	187,2	0,61	15,93
LiZnCoF6	ZnF2	CoF3	1148	17,61	183	0,74	15,69	1230	27	187,2	0,61	15,93
LiCoGaF6	CoF2	GaF3	1400	19,59	159,1	0,745	16,44	1225	28	255	0,62	15,93
LiNiGaF6	NiF2	GaF3	1430	17,6	157,2	0,69	15,31	1225	28	255	0,62	15,93
LiCuRhF6	CuF2	RhF3	1043	16,4	128,5	0,73	16,8	1460	26	175	0,665	15,93
LiZnRhF6	ZnF2	RhF3	1148	17,61	183	0,74	15,69	1460	26	175	0,665	15,93
LiMgVF6	MgF2	VF3	1536	13,68	268,7	0,72	14,72	1679	23,1	271	0,64	21,62
LiFeCrF6	FeF2	CrF3	1375	20,79	158	0,78	16,28	1677	22,5	277	0,615	18,82
		h2	1352,8	16,96	184,9	0,722	15,60	1474	24,7	245,4	0,63	17,45

On the fourth stage we single out in the SPMK the set of the nodes corresponding to the primary objects for compounds of groups G_1 and G_2, and the set of the nodes corresponding to the primary predicates to which arrows from these primary objects approach. The primary objects are represented in the 2-nd and 3-rd columns in tables 1 and 2. The primary object LiF is included into all compounds, therefore its primary properties do not influence belonging of the compound to a certain group and consequently it is not taken into further consideration. On the fifth stage we find distributions of primary predicate labels for G_1 and G_2 groups of compounds. Each of primary variable predicates takes countable set of values - predicate constants. As their labels we took numerical values of properties of the primary objects, which correspond to them. In the example following 5 primary variable predicates were considered: Tm - melting point, So - standard entropy for corresponding simple oxides, H - standard enthalpy formations for corresponding simple oxides, Rs - radius of the ions, C - isobaric thermal capacity. Their values are represented in columns 4-8 for the primary element of the 2-nd column and in columns 9-13 for the primary element of the 3-rd column, thus the labels of predicates are chosen in the way that they coincide with these values.

On the sixth stage typical distributions of labels h_1 and h_2 are calculated through finding the average for the values in columns 4-13 of every table. h_1 and h_2 are represented in the last lines of tab. 1 and 2. Further there are centralized distributions of labels by way of subtraction of the obtained average values of each column from the actual values of their cells. As a result in the numerical cells of both tables we will obtain the values of the transposed matrixes $A^{(1)T}$ and $A^{(2)T}$. On the seventh stage of development matrixes $B^{(1)} = A^{(1)}A^{(1)^T}$ $B^{(2)} = A^{(2)}A^{(2)^T}$ $B = B^{(1)} + B^{(2)}$, an inverse matrix B^{-1}, and matrixes $H = (h^{(1)} - h^{(2)})^T (h^{(1)} - h^{(2)})$ and $H' = H * \frac{KL}{K+L}$, where $K = 10$, $L = 17$ are calculated. On the eighth stage we find an eigen vector W for a nonzero eigen value of the matrix $B^{-1}H'$ (for example, with Matlab software). For the example considered an eigen vector is $W = (c_1, c_2, \cdots, c_{10}) =$ (-0.0002 0.0359 0.0017 0.6283 -0.0276 0.0004 0.0363 0.0015 -0.7754 -0.0242). Cosines of the angles formed by the optimal projective line c with quadrantal angles are proportional to the values of this vector; thus, the coefficient of proportionality does not play any part. On the ninth stage, projections of typical distributions of labels on this line are found: $\bar{z}^{(1)} = W * h_1^T$ and $\bar{z}^{(2)} = W * h_2^T$. We have $\bar{z}^{(1)} = 1.889$, $\bar{z}^{(2)} = 1.6201$. Their common center is $\bar{z} = 0.5(\bar{z}^{(1)} - \bar{z}^{(2)}) = 1.7545$. On the tenth stage we select in the SPMK primary objects for projected

compound in the following way. The objects are selected with connections to the primary predicates, with which the primary objects of group of compounds G_1 also have connections, and with no connections to the primary predicates, with which the primary objects of group of compounds G_2 have connections, thus possible restrictions on the structure of compounds are taken into consideration. Let us assume that the compounds represented in table 3 have been selected.

Table 3

LiMgInF6	MgF2	InF3	1536	13,68	268,7	0,72	14,72	1445	33,5	250	0,8	15,93
LiMnFeF6	MnF2	FeF3	1133	22,25	202,4	0,83	16,24	1300	25	239	0,645	15,93
LiMnGaF6	MnF2	GaF3	1133	22,25	202,4	0,83	16,24	1225	28	255	0,62	15,93
LiMnInF6	MnF2	InF3	1133	22,25	202,4	0,83	16,24	1445	33,5	250	0,8	15,93
LiZnInF6	ZnF2	InF3	1148	17,61	183	0,74	15,69	1445	33,5	250	0,8	15,93
LiCdInF6	CdF2	InF3	1345	20	167,4	0,95	15,93	1445	33,5	250	0,8	15,93
LiMgTiF6	MgF2	TiF3	1536	13,68	268,7	0,72	14,72	1500	21,1	342,2	0,67	15,93
LiMgFeF6	MgF2	FeF3	1536	13,68	268,7	0,72	14,72	1300	25	239	0,645	15,93
LiMgCoF6	MgF2	CoF3	1536	13,68	268,7	0,72	14,72	1230	27	187,2	0,61	15,93
LiFeTiF6	FeF2	TiF3	1375	20,79	158	0,78	16,28	1500	21,1	342,2	0,67	15,93
LiCoTiF6	CoF2	TiF3	1400	19,59	159,1	0,745	16,44	1500	21,1	342,2	0,67	15,93
LiZnTiF6	ZnF2	TiF3	1148	17,61	183	0,74	15,69	1500	21,1	342,2	0,67	15,93
LiZnVF6	ZnF2	VF3	1148	17,61	183	0,74	15,69	1679	23,18	271	0,64	21,62
LiNiCrF6	NiF2	CrF3	1430	17,6	157,2	0,69	15,31	1677	22,5	277	0,615	18,82
LiFeFeF6	FeF2	FeF3	1375	20,79	158	0,78	16,28	1300	25	239	0,645	15,93
LiCoFeF6	CoF2	FeF3	1400	19,59	159,1	0,745	16,44	1300	25	239	0,645	15,93
LiCuFeF6	CuF2	FeF3	1043	16,4	128,5	0,73	16,8	1300	25	239	0,645	15,93
LiZnFeF6	ZnF2	FeF3	1148	17,61	183	0,74	15,69	1300	25	239	0,645	15,93
LiCuCoF6	CuF2	CoF3	1043	16,4	128,5	0,73	16,8	1230	27	187,2	0,61	15,93
LiCoRhF6	CoF2	RhF3	1400	19,59	159,1	0,745	16,44	1460	26	175	0,665	15,93
LiNiRhF6	NiF2	RhF3	1430	17,6	157,2	0,69	15,31	1460	26	175	0,665	15,93
LiCuGaF6	CuF2	GaF3	1043	16,4	128,5	0,73	16,8	1225	28	255	0,62	15,93

To check the correctness of selection, the projection $\mathrm{z} = c_1 x_1^{(3)} + c_2 x_2^{(3)} + \cdots + c_N x_N^{(3)}$ of distribution of labels for each selected connection to the projective straight line is calculated. If $\left|\bar{z}^{(1)} - z\right| < \left|\bar{z}^{(2)} - z\right|$, than the selection is considered the correct one. For compounds of table 3 from top to down in succession we find: z equally 1.8395, 1.9060, 2.0089, 2.1339, 1.8838, 2.0870, 1.6272, 1.6115, 1.5784, 1.7448, 1.6329, 1.6715, 1.6521, 1.6135, 1.7291, 1.6173, 1.5107, 1.6558, 1.4776, 1.5440, 1.4934, 1.6136. As $\bar{z}^{(1)} = 1.889$, $\bar{z}^{(2)} = 1.6201$ only the first 6 compounds according to this technique were selected correctly, and others erroneously. The example considered allows also to check correctness of the technique itself as lattice structure of the compounds of table 3 is known from the beginning: the first 6 compounds have structure of Na2SiF6 crystal lattice type, and all subsequent chemical compounds have structure of Trirutile crystal lattice type. Thus, we receive 100 % of correct answers that proves the technique while the work [4] obtains 86,4 % correct answers for the same group of chemical compounds.

Conclusion

The work considers complex objects with internal structure. A structural - predicate model of knowledge, which is a generalization of the structural - attributive model of knowledge is proposed. A method of obtaining of new knowledge on the structure of complex objects with required properties based on measurement of distances between the groups of the predicates with some interpretation of them is developed in the work. An optimal measure for measurement of these distances with the maximal discernibility of different groups of predicates is constructed. The stages of solution of the problem of complex object development are considered.

Bibliography

1. V.N. Koval, Yu.V. Kuk. Distances between predicates in by-analogy reasoning systems, “Information Theories and Applications”, International Journal, vol. 10, N 1, p. 15-22, Sofia, 2003.
2. V.N. Koval, Yu.V. Kuk. Finding Unknown Rules of an Environment by Intelligent Goal-Oriented Systems, “Information Theories and Applications”, International Journal, vol. 17, N 3, p. 127-138, Sofia, 2001.
3. Гладун В.П. Партнерство с компьютером. Человеко-машинные целеустремленные системы. – Киев: «Port-Royal», 2000. –128 с.
4. Величко В.Ю. Розв’язання дослідницьких задач в дискретних середовищах методами виведення за аналогією. – Киев: Кандидатская диссертация. – 2003. – 150 с.

Authors' Information

Valeriy Koval – The Institute of Cybernetics, Head of Department, address: 40, Prospect Glushkova, Kiev, Ukraine; 03680 e-mail: icdepval@ln.ua

Yuriy Kuk – The Institute of Cybernetics, senior scientific researcher, 40, Prospect Glushkova, Kiev, Ukraine; 03680; e-mail: vkyk@svitonline.com .

CLUSTER MANAGEMENT PROCESSES ORGANIZATION AND HANDLING

Valeriy Koval, Sergey Ryabchun, Volodymyr Savyak, Anatoliy Yakuba

Abstract: The paper describes cluster management software and hardware of SCIT supercomputer clusters built in Glushkov Institute of Cybernetics NAS of Ukraine. The paper shows the performance results received on systems that were built and the specific means used to fulfil the goal of performance increase. It should be useful for those scientists and engineers that are practically engaged in a cluster supercomputer systems design, integration and services.

Keywords: cluster, computer system management, computer architecture.

ACM Classification Keywords: C.1.4 Parallel Architectures; C.2.4 Distributed systems; D.4.7 Organization and Design

1. Cluster Complex Architecture

Basis cluster architecture is the array of servers (contains computing nodes and the control node), are connected among themselves by several local computer networks - a high-speed network of data exchange between computing nodes, a network of dynamic management of a server array and a network for cluster nodes monitoring. User access to cluster as a whole can cope by the access server - a gateway on which check of the rights of access of users to cluster and preliminary preparation of tasks for execution is realized. File services are given user tasks by a file server through the cluster control node. A file server in a system provides data access on file level protocols, like Network File System (NFS). A file server is connected directly to a local data network via high throughput channel. In some cases, the gateway and/or file server functions may be carried out on the control node.

Cluster computing node is a server, more often dual-processor, for direct execution of one user task in one-program mode. Computing nodes are dynamically united through a network in a resource for a specific task, simultaneously on cluster some problems may be executed, depending on amount of free computing nodes.