Chapter 9
Application of Machine Training Methods to Design of New Inorganic Compounds

Nadezhda Kiselyova
A. A. Baikov Institute of Metallurgy and Materials Science of Russian Academy of Sciences, Russia

Andrey Stolyarenko
A. A. Baikov Institute of Metallurgy and Materials Science of Russian Academy of Sciences, Russia

Vladimir Ryazanov
A. A. Dorodnicyn Computing Centre of Russian Academy of Sciences, Russia

Oleg Sen'ko
A. A. Dorodnicyn Computing Centre of Russian Academy of Sciences, Russia

Alexandr Dokukin
A. A. Dorodnicyn Computing Centre of Russian Academy of Sciences, Russia

ABSTRACT

The review of applications of machine training methods to inorganic chemistry and materials science is presented. The possibility of searching for classification regularities in large arrays of chemical information with the use of precedent-based recognition methods is discussed. The system for computer-assisted design of inorganic compounds, with an integrated complex of databases for the properties of inorganic substances and materials, a subsystem for the analysis of data, based on computer training (including symbolic pattern recognition methods), a knowledge base, a predictions base, and a managing subsystem, has been developed. In many instances, the employment of the developed system makes it possible to predict new inorganic compounds and estimate various properties of those without experimental synthesis. The results of application of this information-analytical system to the computer-assisted design of inorganic compounds promising for the search for new materials for electronics are presented.

DOI: 10.4018/978-1-4666-1900-5.ch009

INTRODUCTION

The problem of predicting new multi-component compounds' formation and calculating their intrinsic properties proceeding from the knowledge of their constituent components' properties is one of the most important tasks of inorganic chemistry. Any successful attempt of designing not yet synthesized compounds is of the large theoretical and practical importance. Calculations or predictions, based on only the properties of constituent components, are called *a priori* calculations or predictions. The difficulties of *a priori* predictions are connected with the solution of mathematical problems arising in the quantum mechanical calculations of multi-electronic systems (Gribov, 2010; Kohanoff, 2006). As a result, chemists and materials scientists make use of many empirical prediction methods. It should be noted that inorganic chemistry similar to other empirical sciences, for which, at the modern level of computational mathematics' development, even complex algebraic approaches do not guarantee satisfactory computational results for their objects and phenomena, has various classification schemes since obtaining any scientific knowledge requires two initial stages: data acquisition and data classification. In most empirical sciences, classification schemes play the role of exact mathematical regularities. The development of classification rules is a complicated and laborious process that requires high qualifications of specialists. The application of pattern recognition methods and appropriate software systems allows one to facilitate and speed up the development of classification rules. The tasks of a specialist in a specific subject field when implementing this process are the following: the statement of a problem, choice of objects and phenomena for computer-aided analysis, choice of attribute description, interpretation of results, and application of the classification principles to prediction.

The present chapter is devoted to the use of precedent-based computer training methods for searching for classification rules for inorganic substances and the application of these rules to predicting new compounds and evaluating their properties.

STATEMENT OF THE PROBLEM OF DESIGNING NEW INORGANIC COMPOUNDS

The problem of designing new inorganic compounds can be formulated as the search for combination of chemical elements and their ratio (i.e., determining qualitative and quantitative compositions) for the synthesis (under given conditions) of the predefined space molecular or crystal structure of a compound that possesses the required functional properties. It is the knowledge of the properties of chemical elements and data about other compounds already investigated that constitute initial information for the calculations. The problem of designing new inorganic compounds can be reduced to discovering the relationships between the properties of chemical systems (for example, properties of inorganic compounds) and the properties of elements that form these systems (Burkhanov & Kiselyova, 2009; Kiselyova, 2005).

The methods of pattern recognition are one of the most effective means of search for regularities in the large arrays of chemical data. In this case, the problem can be defined as follows (Zhuravlev, Kiselyova, Ryazanov, Senko, & Dokukin, 2011). Suppose that every inorganic substance is described by a vector $\mathbf{x} = (x_1^{(1)}, x_2^{(1)},..x_M^{(1)}, x_1^{(2)}, x_2^{(2)},.., x_M^{(2)},\ldots, x_1^{(L)}, x_2^{(L)},.., x_M^{(L)})$, where L is the number of chemical elements that form a compound and M is the number of parameters of chemical elements. Each substance is also characterized by a class membership parameter: $a(x) \in \{1, 2,\ldots, K\}$, where K is the number of classes. The training sample consists of N objects: $\mathbf{S} = \{\mathbf{x}_i, i = 1, \ldots, N\}$. We denote a subset of objects of the training sample from class a_j, j = 1, 2, …, K,

by $\mathbf{S}_{aj} = \{x: a(x) = a_j\}$. The aim of training is to construct a classification rule that distinguishes not only between objects of different classes in the training sample but also preserve prognostic ability to generate new combinations of chemical elements that were not used for training. Thus, we deal with the classical statement of a precedent-based pattern recognition problem. The peculiarity of the subject field manifests itself only via the formation of attribute description possessing a composite structure: the set of parameters of chemical elements (the components of an inorganic substance) is repeated as many times as there are elements included into the compound.

METHODS AND TOOLS

The main difficulties of pattern recognition application to solving tasks of inorganic chemistry are the following: small informative gain of attributes - properties of chemical elements, the strong correlation of these attributes owing to their dependence on common parameter - atomic number of chemical elements (it follows from the Periodic Law), blanks of attributes' values; in many cases, we have the large asymmetry of training set size for different classes; sometimes attribute description includes non-numerical attributes (symbolic), and there are experimental mistakes of classification in training sets.

In connection with the above-stated peculiarities of subject domain, the search for methods and algorithms of pattern recognition allowing the correct solution of these problems was one of the basic tasks at computer-assisted design of inorganic compounds. It was established during testing various algorithms of machine training that it is impossible to specify beforehand, what algorithm is most effective in solving a certain chemical task of inorganic compounds design. Quite often programs, which have classified training set well, give bad results at the prediction of unknown compounds. In this connection, the most effective way of solving tasks of predicting new inorganic compounds and their properties is concerned with the methods of recognition by ensembles of algorithms (Zhuravlev, Ryazanov, & Sen'ko, 2006). At synthesis of a collective decision it is possible to compensate mistakes of separate algorithms by the correct predictions of other algorithms.

Another way of increasing the accuracy of predicting is the use of chemical element properties' dependence on atomic number. On the one hand, attribute descriptions are formed from parameters of elements with strong mutual correlation. This fact complicates searching for properties that are the most important for classification. On the other hand, the classifying regularities including values of any subset of properties of chemical elements used for the description of inorganic compounds should, in principle, give identical results of classification. I.e., the results of the prediction with use of various subsets of properties of elements should, basically, coincide. This fact allows an additional possibility of collective decision making but already on the basis of some sets of attribute descriptions obtained as a result of division of an initial set of properties of chemical elements on partially crossed subsets.

The problem of filling blanks of attribute values also is partially solved with the use of periodic dependences of elements' parameters (Kiselyova, Stolyarenko, Ryazanov, & Podbel'skii, 2008; Kiselyova et al., 2011). A skipped parameter value of some element is replaced by the average value of this parameter for two chemical elements that are nearest (within the range of group of Periodic System) to the element in question. In this case, the average value of involved property over the nearest elements is computed. Here the relative Euclidean distance between elements need to be not greater than 10%, and these elements are sought only among the elements in the same group of the Periodic System. If no appropriate element is found, then either the blank is replaced by the mean value of the element's property for the substances

with the same classifying attributes (in case of the training sample), or this attribute is excluded from the sample (in case of the control sample for recognition), and the system is retrained again without this property, i.e., the resulting sample after eliminating this parameter is passed to the input of a pattern recognition procedure.

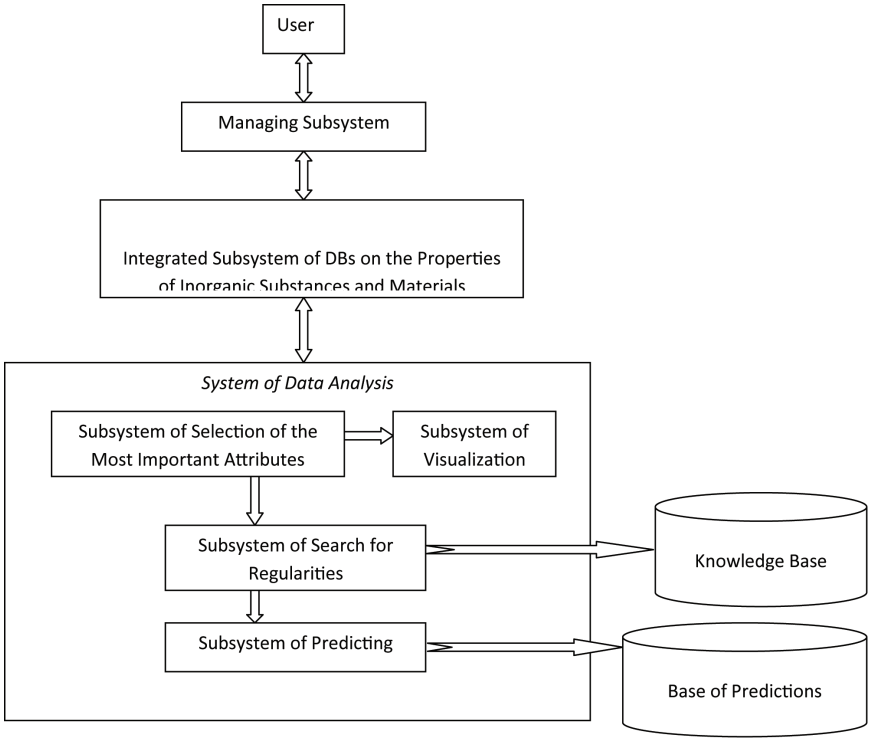
After testing some pattern recognition program systems, we have chosen a wide class of algorithms of the system RECOGNITION developed at A. A. Dorodnicyn Computing Centre, Russian Academy of Sciences (Zhuravlev et al., 2006). This multifunctional system for pattern recognition includes:

1. Standard statistical methods:
   a. The *k*–nearest neighbors,
   b. Fisher's linear discriminator.
2. Linear machine method implementing linear solving rule that is calculated by searching for maximal subsets of simultaneous inequalities corresponding to recognized objects. This maximal subset is searched with the help of relaxation method.
3. Neural networks. Besides standard variant of multilayer perceptron, its modification with three layers is used. With that the output of the second layer includes a variety of products of the first layer outputs.
4. Support vector machine.
5. Estimate calculating algorithms (EC) model, where the estimates for classes are calculated as weighted sum of similarity functions by a variety of features subsets and standard objects.
6. LoReg (Logical Regularities) voting algorithm where the estimations for classes are calculated with the help of voting by a logical regularities' system. Logical regularity is defined as a non-extendible multidimensional interval in features space that includes objects from one of recognized classes (or mainly from one class).
7. Deadlock test algorithm, where the estimates for classes are calculated as sums of similarity functions by a variety of deadlock tests and standard objects. Deadlock test is defined as irreducible combination of features that allows separation of objects from different classes.
8. Statistical weighted syndromes, where the estimates for classes are calculated by systems of so called «syndromes». Syndrome is defined as a sub-region in feature space where the fraction of one of recognized classes significantly differs from the fraction of the same classes in neighbor sub-regions. Syndromes are searched with the help of optimal partitioning technique.
9. Decision trees.
10. Collective methods where final solution is calculated by set of previously trained algorithms belonging to above mentioned families (algebraic, logical, and heuristic correctors).
11. The set of unsupervised classification methods with constructing collective solutions.

Also the program of concept formation ConFor developed at V. M. Glushkov Institute of Cybernetics, National Academy of Sciences of Ukraine (Gladun, 1997, 1995, 1972; Gladun & Vashchenko, 1975) was used with success. The system is based on a special data structure named the growing pyramidal networks.

The special information-analytical system (IAS) for design of inorganic compounds was developed (Burkhanov & Kiselyova, 2009; Kiselyova et al., 2011). Apart from subsystem of data analysis based on above mentioned algorithms of pattern recognition, IAS includes (Figure 1) an integrated subsystem of DBs for the properties of inorganic substances and materials, subsystem of selecting the most important attributes, visualization subsystem, knowledge base, base of predictions for various classes of inorganic substances, and managing subsystem.

Figure 1. Schema of information-analytical system for design of inorganic compounds



The subsystem of databases on the properties of inorganic substances and materials developed at the A. A. Baikov Institute for Metallurgy and Materials Science, Russian Academy of Sciences (IMET RAS), is the source of information for the computer-aided analysis. Its usage allows a formation of representative training sample. Now this subsystem incorporates the following DBs:

1. The DB "Phases" for the properties of inorganic compounds (Kiselyova, 2005; Kiselyova, Dudarev, & Zemskov, 2010) now containing information on more than 49 000 ternary compounds (i.e., compounds made up of three chemical elements) and more than 23 000 quaternary compounds, extracted from more than 26 000 publications.
2. The DB "Elements" for the properties of chemical elements (Kiselyova et al., 2010) containing data for more than 90 parameters.
3. The DB "Diagrams" for phase diagrams of semiconductor systems (Kiselyova, 2005; Kiselyova et al., 2010, 2004) containing the information on phase diagrams of semiconductor systems and the physical and chemical properties of phases forming in those, collected and evaluated by experts. At present this DB comprises the detailed information on several tens of systems that are the most important in semiconductor electronics.
4. The DB "Crystal" for the properties of acousto-optical, electro-optical, and nonlinear optical substances (Kiselyova, 2005; Kiselyova et al., 2010, 2004) now containing the information on the parameters of more than 140 materials.
5. The DB "Bandgap" for the forbidden band width of inorganic substances (Kiselyova et al., 2010) currently containing the information on more than 3000 substances.

The total size of DBs is about 8 Gbytes. The integrated subsystem of DBs allows specialists to gain aggregate information on the properties of substances and materials from different databases at one time. Authorized users can access to the subsystem of DBs via the Internet (http://imet-db.ru).

The subsystem of selecting the most important attributes (properties of chemical elements) is based on a procedure of minimization of generalized error functionals for convex correcting procedures with respect to ensembles of predictors constructed on the basis of individual attributes (Senko, 2009; Senko & Dokukin, 2010). The selection of the properties of chemical elements, providing the most information for the classification of substances, is of double significance. On the one hand, it enables a drastic reduction of attribute description that includes hundreds of elements' properties for multi-component substances. On the other hand, the selection of the most important properties of elements in classification of chemical substances affords physical interpretation of the resulting classifying regularities improving the confidence in the obtained predictions and makes it possible to find substantial causal relationships among the parameters of subjects and to develop physical and chemical models of the phenomena.

The visualization subsystem allows an illustration of attribute selection results. The information about properties of chemical compounds is represented in habitual for the chemists and materials scientists form: as projections of the points corresponding to compounds of a certain type in properties space of chemical elements. The system of visualization is intended for the representation of information about the coordinates of properties for elements included into compounds. Use of algebraic functions of chemical elements' properties is possible, for what the special complex attributes formation subsystem was developed.

At the solution of recognition tasks and data analysis, the tools for visualizing multidimensional data are important. They allow a graphic representation of a configuration of classes, clusters, and disposition of separate objects - chemical compounds. These tools are necessary, first of all, in case of tasks with the large number of attributes, when the separate projections in 2D- or 3D-subspaces of attributes contain poor information concerning n-dimensional descriptions. In this connection, the subsystem of multidimensional scaling was developed too.

The knowledge base (tasks base) contains the discovered regularities, which can be used for prediction of substances not yet synthesized and estimation of their properties when there is no information on a certain chemical system in the databases. The regularities are stored in the tasks base in the intrinsic format of those software products for the data analysis by whose means they were obtained. Such implementation makes it possible to integrate new software products for the analysis of data into the IAS and resolves the problem associated with the fact that the forms of representation of the resulting regularities in the computer training methods used are substantially different. By a task, it is meant the procedure of training by the selected methods on a particular training sample. Here it is suggested that not the results of training, as such (like logical expressions or the structure of a trained neural network), but so-called labels for the tasks be stored in the tasks base. The term label is taken to mean the necessary information for the task, which permits distinguishing this task from others. The following information on the task is stored in the IAS: the unique number of the task; the training sample in standard format; data for the attributes used to form the training sample; the identifier of the software product for the data analysis by whose means the regularities were obtained; the list of methods employed in training, with their parameters; information on the quantitative and qualitative composition of the compounds used in training; the identifier of the compounds' parameter to be predicted; etc.

The predictions base contains the results of previous computer experiments, as well as references to service information stored in the knowledge base. The use of the predictions base made it possible to improve the functionality of the IMET RAS DB for the properties of inorganic substances and materials through providing the user not only with the available information on substances that have already been studied, but also with predictions of inorganic compounds which have not yet been obtained and estimations of their properties.

The managing subsystem organizes the computing process and carries out interaction among all functional subsystems of the IAS, as well as provides access to IAS via the Internet. Besides, the managing subsystem provides the user with software for preparation of data for the analysis, produces reports in the form habitual for chemists, and provides other service functions. In particular, a special subsystem has been designed to retrieve the DB information that, after its estimation by chemist, is used to train the computer, and to prepare this information for the subsequent analysis. The subsystem allows the chemist to edit found information and to form an attribute description of compound, which is a complex description made up of parameters of few chemical elements included into its composition. The chemist selects the properties of chemical elements to form a training sample, and the subsystem for preparation of training set retrieves the chosen values of the elements' properties from the DB "Elements", makes up complex attributes as algebraic functions of the initial parameters of elements when needed, and merges the attribute description to produce a table that is there upon passed to the input of the prediction subsystem. The subsystem for generation of results is intended for presenting predictions in the tabular form customary among chemists and material scientists.

Now the IAS is the main tool for predicting new inorganic compounds in our investigations.

APPLICATION OF PATTERN RECOGNITION METHODS TO INORGANIC CHEMISTRY AND MATERIALS SCIENCE

The first studies of the application of the pattern recognition methods for predicting inorganic compounds were carried out in IMET by Savitskii and his co-workers in the mid-1960s (Savitskii, Devingtal', & Gribulya, 1968a, 1968b). The results of the prediction for rather simple binary metallic systems turned out to be excellent: their comparison with new obtained experimental data showed that the reliability of the prediction of the binary compounds exceeded 90% (Savitskii & Gribulya, 1985; Savitskii et al., 1990). Software developed by Devingtal' (Devingtal', 1971, 1968) was used in these pioneer studies on designing binary intermetallics. Devingtal' applied the methods of linear programming to the solution of the extreme problems of pattern recognition. The successful investigations of Savitskii's team initiated the application of pattern recognition methods to solution of various tasks in physical metallurgy. In (Vozdvizhenskii, 1974; Vozdvizhenskii & Falevich, 1973), an algorithm of potential functions (Iserman, Braverman, & Rosonoer, 1970) was used for the prediction of binary metallic systems. The estimate calculation algorithm (Zhuravlev & Nikiforov, 1971) was applied to the search for optimal quantities of additives of different chemical elements to the steel for attaining the extreme mechanical properties (Lenovich, 1974). A linear algorithm of computer training of pattern recognition was applied to the search for alloy dopants (Gulyaev & Pavlenko, 1973). Mechanical properties of steels were predicted (Li, 2006) with the application of the algorithm of binary decision trees (Breiman, Friedman, Olshen, & Stone, 1984) and artificial neural networks (ANN) (Bahrami, Mousavi, & Ekrami, 2005). ANN were used in predicting shear strength of Ni-Ti alloys (Taskin, Dikbas, & Caligulu, 2008) and in predicting the mass

loss quantities of some Al–Cu based composite materials reinforced with SiC (Hayajneh, Hassan, Alrashdan, & Mayyas, 2009).

However most of applications of pattern recognition methods were connected with solution of the problems of inorganic chemistry and materials science. A significant advance has been made in this area by teams working at A. A. Baikov Institute for Metallurgy and Materials Science of RAS (Kiselyova, 2005, 2003, 2002, 1993a, 1993b; Kiselyova, Gladun, & Vashchenko, 1998; Kiselyova, LeClair, Gladun, & Vashchenko, 2000; Kiselyova, Pokrovskii, Komissarova, & Vashchenko, 1977; Kiselyova, Ryazanov, & Sen'ko, 2009; Kiselyova & Savitskii, 1983; Kiselyova, Stolyarenko, Gu, & Lu, 2007a; Kiselyova et al., 2011, 2008, 2007b; Savitskii et al., 1990, 1968a, 1968b; Savitskii & Gribulya, 1985; Savitskii, Gribulya, & Kiselyova 1982, 1981, 1980, 1979; Zhuravlev et al., 2011) and at University of Shanghai (Chen, Chen, Lu, Li, & Villars, 1999; Chen, Li, & Qin, 1998; Chen, Li, Yao, & Wang, 1996a, 1996b; Chen, Lu, Chen, Qin, & Villars, 1999; Chen, Lu, Qin, Chen, & Villars, 1999; Chen, Lu, Yang, & Li, 2004; Chen, Zhu, & Wang, 2000; Gu, Lu, Bao, & Chen, 2006; Liu, Chen, & Chen, 1994; Lu et al., 1999; Yan, Zhan, Qin, & Chen, 1994). Chinese researchers used in their calculations: ANN (Chen et al., 1998, 1996b), genetic algorithms (GA) (Chen et al., 1998), support vector machine (SVM) methods (Chen et al., 2004; Gu et al., 2006), partial least squares (PLS) regression method (Chen et al., 1996a), principal component backing (PCB) method (Liu et al., 1994), etc. The investigators of IMET RAS used methods developed by Devingtal' (Savitskii et al., 1990, 1982, 1981, 1980, 1979, 1968a, 1968b; Savitskii & Gribulya, 1985), various modifications of the program of concept formation (Savitskii et al., 1990, 1982, 1981, 1980, 1979; Kiselyova, 2005, 2003, 2002, 1993a, 1993b; Kiselyova & Savitskii, 1983; Kiselyova et al., 2011, 2009, 2008, 2007a, 2007b, 2000, 1998, 1977), and software product RECOGNITION (Kiselyova et al., 2011, 2010b, 2009, 2008, 2007a, 2007b; Zhuravlev et al., 2011). It should be noted that the algorithms based on symbolic machine training methods (for example, ConFor (Gladun, 1997, 1995, 1972; Gladun & Vashchenko, 1975) or LoReg (Kovshov, Moiseev, & Ryazanov, 2008; Ryazanov, 2007) are the most adequate for solution of the tasks of inorganic compounds design because they provide a possibility of analyzing "mixed" chemical data (numerical, symbolic, etc.). As a rule, these algorithms give good results of predicting new inorganic compounds also. Note should be taken that traditional expert systems (Chen et al., 1998; Zhou, Jin, Shao, & Chen, 1989; Yao, Qin, Chen, & Villars, 200l) are poorly suitable in inorganic chemistry. The keystone at development of expert systems in this area is unsolved problem of knowledge acquisition from specialists.

The various methods of pattern recognition were used by other investigators for designing new inorganic substances with predefined properties. The discriminant analysis was widely used for the prediction of the possibility of formation of binary compounds with lanthanides (Kutolin & Kotyukov, 1978; Kutolin, Vashukov, & Kotyukov, 1978), of the crystal structure type for refractory binary compounds (Kutolin & Kotyukov, 1979), of the type and concentration of defects and of defect formation energetics in imperfect crystals of refractory compounds (Kutolin, Komarova, & Frolov, 1982). The analysis of the electrical properties of PZT ceramics was carried out by a back propagation artificial neural network method (Cai, Xia, Li, & Gui, 2006). The ANN were used in predicting ultra-hard binary compounds (Thaler, 1998), new orthorhombic ABO_3 perovskites (Aleksovska, Dimitrovska, & Kuzmanovski, 2007), the band gap energy and the lattice constant of chalcopyrites (Zeng, Chua, & Wu, 2002), the

hexagonal lattice parameters of apatites (Kockan & Evis, 2010). Principal component analysis (PCA) was used in the search for hydrogen storage for AB_5-alloys (Ye, Xia, Wu, Du, & Zhang, 2002). New electro-ceramic materials were designed using ANN and GA (Scott, Manos, & Coveney, 2008). The method of potential functions was used for predictions of new spinels (Talanov & Frolova, 1979, 1982). The promising approach is connected with combination of machine training techniques and first principles computations (Ceder, Morgan, Fischer, Tibbetts, & Curtarolo (2006); Hautier, Fischer, Jain, Mueller, & Ceder, 2010).

These examples do not exhaust all the applications of pattern recognition methods to inorganic chemistry and materials science. More detailed reviews are given in the monograph (Kiselyova, 2005) and the reviews (Burkhanov & Kiselyova, 2009; Kiselyova, 2002).

RESULTS OF THE PREDICTION OF NEW INORGANIC COMPOUNDS

The potential of computer training methods for pattern recognition in designing new inorganic compounds can be demonstrated by comparison of the results of the predictions with newer experimental data.

The problem of designing new substances with desired properties can be divided into four consecutive problems:

- The prediction of compound formation or non- formation;
- The prediction of compounds of desired composition;
- The prediction of compounds with a specific crystal structure type;
- The estimation of quantitative properties of compounds (critical temperature of transition to superconducting state, homogeneity region, bandgap, etc.).

Prediction of Compounds Formation with Composition ABO_3

The compounds with composition ABO_3 are conventional piezoelectric, acousto-optic, electro-optic and nonlinear optical materials. The prediction of these compounds was the first experience of computer-assisted design of multi-component substances (Kiselyova et al., 1977). At the solution of prediction task of forming compounds with this composition, 239 examples of formation and 39 examples of non-formation of compounds ABO_3 were used for computer training.

Based on physical and chemical understanding of the nature of substances of this kind, three sets of component properties were chosen for description of these substances:

1. The distribution of electrons in the energy levels of isolated atoms of the chemical elements A and B and their formal valences in compounds with this composition.
2. The types of incomplete electronic shells (s, p, d, or f), the first four ionization potentials, the ionic radii according to Bokii and Belov, the standard isobaric heat capacities and the formal valences of the elements A and B in these compounds.
3. The ionic radii according to Bokii and Belov, the standard enthalpies of formation and isobaric heat capacities of appropriate simple oxides, and the formal valences of the elements A and B in these compounds.

The program based on concept formation process (Gladun, 1997, 1995,1972; Gladun & Vashchenko, 1975) was applied to computer training (machine learning). Regularity classifications and the predictions of the formation of unknown compounds with composition ABO_3 were obtained separately for each of the three sets of properties of constituent components (attributes). As a result, we obtained three tables of predictions. Next, we compared the predictions in these three tables and made a decision on the existence of a given compound for which the predictions were not contradictory. The part of table illustrating the prediction of compounds formed by two- and four-valent elements is given in Table 1.

Table 1. Predictions of compounds with composition $A^{II}B^{IV}O_3$

A^{II} B^{IV}	Be	Mg	Ca	Ti	V	Mn	Fe	Co	Ni	Cu	Zn	Ge	Sr	Pd	Cd	Sn	Ba	Hg	Pb	Ra
C	↔	⊕	⊕	©	+	⊕	⊕	⊕	⊕	⊕	⊕	+	⊕	+	⊕	©	⊕	©	⊕	⊕
Si	©	⊕	⊕	+	⊗	⊕	⊕	©	©		⊕	+	⊕	+	⊕	©	⊕	+	⊕	+
S			©	+	+	©	©	©	+		+	+	©	+	©	©	©	©	©	+
Ti	↔	⊕	⊕	+	+	⊕	⊕	⊕	⊕	©	⊕	+	⊕	©	⊕	+	⊕	⊕	⊕	+
V			⊕	+	+	©	+	⊕	⊕	⊕	+	+	©	+	⊕	+	©	©	+	+
Mn	+	©	⊕	+	+	+	©	⊕	©	©	©	+	⊕	+	©	©	⊕	+	©	+
Ge	+	⊕	⊕	+	+	⊕	⊕	©	+	⊕	©	+	⊕	+	⊕	©	⊕	+	⊕	+
Se		⊕	©	+	+	©	©	©	©	©	©	+	©	©	©	+	©	©	©	©
Zr	↔	↔	⊕	+	+	+	↔	+	+	+	©	+	⊕	+	⊕	+	⊕	+	⊕	+
Mo		⊕	©	+	+	©	⊕	©	©	+	©	+	⊕	+	+	+	⊕	+	+	+
Tc			©	+	+	+	+	+	+	+	+	+	©	+	+	+	©	+	©	+
Ru			©	+	+	+	+	+	+	+	+	+	⊕	⊗	+	+	⊕	+	©	+
Sn			⊕	+	+	©	+	⊕	©	+	⊕	+	⊕	+	⊕	+	⊕	©	⊕	+
Te		⊕	⊕	+	+	⊕	+	©	⊕	⊕	⊕	+	⊕	+	©	+	©	©	©	+
Ce	↔	⊕	⊕							⊗	+		⊕	+	⊕		⊕	+	⊕	⊕
Pr										+	+		©	+	+		⊕	+	+	+
Tb	-	-	+							+	+		©	+	+		⊕			
Hf		↔	⊕	+	+	+	+	+	↔	+	+	+	⊕	+	©	©	⊕		⊕	
Ta			+	+	+	+	+	+	+	©	+	+	+	+	+	©	+			
W		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+				
Re	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	©			
Os	-	-	⊕	+	+	+	+	+	+	+	+	+	⊕	+	©	+	⊕			
Ir			⊕	+	+	+	+	+	+	+	+	+	⊕	+	+	+	©			
Pt	-	-	©	+	+	+	+	+	+	+	+	+	+	+	+	+				
Pb			⊕	+	+	©	©	+	+	©	⊕	+	⊕	+	⊕	©	⊕	⊕		
Po			+	+	+	+	+	+	+	+	+	+	©	+	+	+	©			
Th	↔	⊕	⊕	+	+	+	+	+	+	+	+	+	⊕	+	⊕	+	⊕		⊕	
Pa	-	-	+	+	+	+	+	+	+	+	+	+		+	+	+				
U	↔	↔	©	+	+	+	+	©	+	+	+	+	©	+	⊕	+	⊕			

In the last three decades 90 predictions were tested experimentally. Only 3 predictions of compounds with compositions $VSiO_3$, $CuCeO_3$ and $PdRuO_3$ were erroneous, i.e., the error of prediction was equal to 3%.

Prediction of the New Langbeinites with Composition $A_2B_2(XO_4)_3$

The more complicated compounds with composition $A_2B_2(XO_4)_3$ were designed (Kiselyova et al., 2000) using pattern recognition method (Gladun, 1997, 1995, 1972; Gladun & Vashchenko, 1975). The compounds with this composition and langbeinite crystal structure type belong to promising class of piezoelectric, ferroelectric, electro-optic, nonlinear optical, and luminescent substances.

The substances were classified into four classes: (1) compounds with composition $A_2B_2(XO_4)_3$ and langbeinite crystal structure type (29 examples); (2) compounds with composition $A_2B_2(XO_4)_3$ and $K_2Zn_2(MoO_4)_3$ crystal structure type (7 examples); (3) compounds with this composition and a crystal structure different from those listed above (5 examples), (4) non-formation of compounds with composition $A_2B_2(XO_4)_3$ under ambient conditions (26 examples).

Designations: + formation of compound with composition ABO_3 is predicted; - formation of compound with composition ABO_3 is not predicted; ⊕ compound with composition ABO_3 was synthesized and appropriate information was used in the computer training process; ↔ compound with composition ABO_3 does not exist under normal conditions and this information was used in the computer training process; © predicted formation of compound with composition ABO_3 which was confirmed by experiment; ⊗ predicted formation of compound with composition ABO_3 which was not confirmed by experiment. Here and in other Tables the blank spaces correspond to the disagreement of the predictions with the use of different attribute sets.

Three sets of attributes were used for description of compounds:

1. The distribution of electrons in the energy levels of isolated atoms of the chemical elements A, B and X and the ionic radii according to Shannon of ions A^+ (C.N. = 12), B^{2+} (C.N. = 6), and X^{6+} (C.N. = 4).
2. The first three ionization potentials, the above-mentioned ionic radii according to Shannon, the electronegativities according to Pauling, the standard isobaric heat capacities, and the standard entropies of individual substance, Debye temperatures, energies of crystal lattice, the melting and boiling points, heats of melting and boiling, and the relations of atomic number to atomic weight for elements A, B, and X.
3. The melting points, the standard enthalpies of formation, isobaric heat capacities and Gibbs energies of appropriate simple oxides A_2O, BO и XO_3, and the ionic radii according to Shannon of ions in these oxides.

The results of comparison of predictions using above-mentioned attributes sets are shown in Table 2. In the last decade, 17 predictions were tested experimentally. In five cases the results turned out to be incorrect, i.e., the prediction error for these complicated compounds was 29%.

Designations: L formation of compound with the langbeinite crystal structure type is predicted; K - formation of compound with the crystal structure type $K_2Zn_2(MoO_4)_3$ is predicted; - the crystal structure differing from those listed above is predicted; L, K compound with corresponding type of crystal structure was synthesized and appropriate information was used in the computer training process; ↔ compound with the crystal structure differing from those listed above does not exist under normal conditions and this information was used in the computer training process; (*) compound $A_2B_2(XO_4)_3$ is not formed and this fact was used in the computer learning process;

Table 2. Part of predictions of the crystal structure type for compounds with composition $A_2B_2(XO_4)_3$

X	S					Cr					Mo					W				
A B	Na	K	Rb	Cs	Tl	Na	K	Rb	Cs	Tl	Na	K	Rb	Cs	Tl	Na	K	Rb	Cs	Tl
Mg	Lў	(L)	(L)	(*)	L	L	L©		L©	L©	Kў	(K)	(L)	(L)	(L)		↔	(L)	L©	L
Ca	(*)	(L)	L©	(L)	(*)			L	L	L	(*)	?	?	?	?	(*)	*	?	?	?
Mn	(*)	(L)	(L)	L	(L)	L		(L)	L©	L	K	↔	(L)	(L)	(L)					
Fe	*	L©	L©		(L)	L	K	L	L	L	K	K	?	?	?		K			
Co	(*)	(L)	L©		(L)	L	K	L	L	L		(K)	(L)	(L)	↔		K			
Ni	(*)	(L)	L©	L	L	L		L	L	L	K	(K)	(L)	(L)	(L)					
Cu	(*)		L	*	L	L	K	L	L	L	K	(K)	?	?	?		K			
Zn	*	(L)	L	*	L	L	K	L	L	L	↔	(K)	(K)	-	(K)		Kў			
Sr	(*)	?	?		(*)	*	*	?	?	?	(*)	?	?	?	?	(*)	*	*	*	*
Cd	(*)	(L)	(L)		(L)							K	↔	(L)	Kў	(*)		L	L	L
Ba	(*)		(*)	(*)	(*)	*	*				(*)		*	*	*	*	*			
Pb					*	(*)	*	*©	*	*	(*)	*©	(*)	*ў	*	(*)	*	(*)	(*)	*

* formation of compound with composition $A_2B_2(XO_4)_3$ is not predicted; © prediction was confirmed experimentally; ў prediction was not confirmed experimentally; here and in other Tables ? corresponds to the indefinite result.

Prediction of the New Intermetallic Compounds with Composition AB_2Si_2

The intermetallics with $ThCr_2Si_2$ crystal structure type are investigated intensively because of their ferro- and anti-ferromagnetic properties. We predicted the new compounds with this crystal structure and with composition AB_2Si_2 (Kiselyova & Savitskii, 1983) using software (Gladun, 1997, 1995, 1972; Gladun & Vashchenko, 1975).

Each substance was represented in the computer memory as a set of especially coded values (Gladun, 1997, 1995, 1972; Gladun & Vashchenko, 1975) of the properties of elements A and B, whose class ((1) a compound with crystal structure type $ThCr_2Si_2$ and (2) a compound with the crystal structure differing from $ThCr_2Si_2$ or non-formation of compound with composition AB_2Si_2 under normal conditions) is indicated as the target feature. The searches for classifying regularities and predictions were carried out using two sets of properties of elements A and B:

1. The distribution of electrons in the energy levels of isolated atoms of the chemical elements A and B.
2. The first three ionization potentials, the metal radii according to Bokii and Belov, the standard entropies of individual substance, the melting points, the number of complete electronic shells, the number of electrons in incomplete s-, p-, d-, f-electronic shells for the atoms of elements A and B.

Shown in Table 3 are some of the predictions of new compounds of this type. An experimental check showed that out of 120 predictions checked only fifteen were wrong (the prediction error is 13%).

Designations: +)formation of compound with the crystal structure type $ThCr_2Si_2$ is predicted; - formation of compound with the crystal structure type $ThCr_2Si_2$ is not predicted; ⊕ compound with

Table 3. Part of predictions of the $ThCr_2Si_2$ crystal structure type for compounds with composition AB_2Si_2

BA	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ru	Rh	Pd	Ag	Os	Ir	Pt	Au
Ca	+	⊕	↔	⊕	⊕	⊕	⊕	+	+	©	⊕	+	+	+	⊕
Sr		⊕		⊕	⊕	⊕	⊕	-	-		⊕			©	⊕
Y	⊕	⊕	⊕	⊕	⊕		-				-	+	©	+	⊕
Zr	+	+	⊕	⊕	⊕						-	+	+		
Ba	+	⊕	+	⊕	⊕	⊕	⊕	+	⊕	⊕	⊕	+	⊕	⊕	⊕
La	+	⊕	⊕	⊕	⊕	⊕	⊕	⊕	©	©	⊕	+	©	⊕	⊕
Ce	↔	⊕	⊕	⊕	⊕	⊕	⊕	⊕	©	⊕	⊕	©	©	⊕	⊕
Pr	+	⊕	⊕	⊕	⊕	©	⊕	©	+	©	⊕	©	©	+	⊕
Nd	©	⊕	⊕	⊕	⊕	©	⊕	⊕	©	©	⊕	+	©	⊕	⊕
Pm	+	+	+	+	⊕	+	+	+	+	+	+	+	+	+	+
Sm	©	⊕	⊕	⊕	⊕	©	+	⊕	+	©	⊕	+	©	⊕	⊕
Eu	+	+	⊕	⊕	⊕	⊕	©	+	©	©	⊕	+	©	⊕	⊕
Gd	©	⊕	⊕	⊕	⊕	©	+	©	©	⊕	⊕	©	⊕	⊕	⊕
Tb	©	⊕	⊕	⊕	⊕	©	+	⊕	©	©	+	©	©	+	⊕
Dy	©	⊕	⊕	⊕	⊕	©	+	⊕	©	©	©	+	+	⊕	⊕
Ho	©	⊕	⊕	⊕	⊕	©	+	©	+	©	+	©	+	⊕	⊕
Er	©	⊕	⊕	⊕	⊕	©	+	⊕	©	©	+	©	©	⊕	⊕
Tm	©	©	⊕	⊕	⊕	©	+	©	+	©	©	+	+	⊕	+
Yb	©	⊕	⊕	⊕	⊕	©	+	⊕	⊕	⊕	⊕	+	+	⊕	⊕
Lu	©	©	⊕	⊕	⊕	©	+	©	©	©	+	+	+	⊕	+
Hf	+	+	↔	⊕	⊕										
Ac	+	+	+	+	+	+		+	+	+	+	+	+	+	+
Th	⊕	⊕	⊕	⊕	⊕	⊕		©	©	©	+	©	©	©	©
Pa	+	+		+	+	+		+	+	+	+	+	+	+	+
U	©	©		©	⊕	©		©	©	©	+	©	⊕	©	©
Np	⊕	⊕	⊕	⊕	⊕			©	©	©	+	©	⊕	⊕	⊕
Pu	©	©		©	©	©		©	©	©	+	©	⊕	⊕	©+
Am	+	+		+	+	+	+	+	+	+	+	+	+	+	+
Cm	+	+		+	+	+		+	+	+	+	+	+	+	+

the crystal structure type $ThCr_2Si_2$ was synthesized and appropriate information was used in the computer training process; ↔ compound with the crystal structure type $ThCr_2Si_2$ does not exist under normal conditions and this fact was used in the computer training process; © - predicted formation of compound with the crystal structure type $ThCr_2Si_2$ which is confirmed by experiment; ⊗ predicted formation of compound with the crystal structure type $ThCr_2Si_2$ which is not confirmed by experiment; ∅ predicted absence of compound with the crystal structure type $ThCr_2Si_2$ which is not confirmed by experiment.

The advantage of computer-assisted design methods is the possibility of fast correction of the classification regularities with the appearance of new compounds for which the experimental information contradicts the obtained predictions. To this end, one should just add new examples to those previously used in the computer analysis

and perform additional training of the information-analytical system. Taking into account that in the past years 120 new compounds with composition AB_2Si_2 were synthesized, it was decided to carry out computer training with inclusion of newer data, with wider combination of properties of the elements and with a set of new pattern recognition procedures.

Recently we predicted the formation and crystal structure type under ambient conditions for compounds with composition AB_2X_2 (X – B, Al, Si, P, Ga, Ge, As, Se, Sn, Sb, or Te). The IAS developed by us was used for computer training.

The substances were classified into seventeen classes: (1) compounds with composition AB_2X_2 and $ThCr_2Si_2$ crystal structure type (649 examples); (2) compounds with $FeMo_2B_2$ structure type (261 examples); (3) compounds with $CaAl_2Si_2$ structure type (133 examples); (4) compounds with $CaBe_2Ge_2$ structure type (86 examples); (5) compounds with $NiMo_2B_2$ structure type (58 examples); (6) compounds with $CoSc_2Si_2$ structure type (47 examples); (7) compounds with ZnK_2O_2 structure type (31 examples); (8) compounds with $CaRh_2B_2$ structure type (24 examples); (9) compounds with $AlMn_2B_2$ structure type (22 examples); (10) compounds with PdK_2P_2 structure type (18 examples); (11) compounds with PtK_2S_2 structure type (12 examples); (12) compounds with $LaPt_2Ge_2$ structure type (12 examples); (13) compounds with SnU_2Pt_2 structure type (11 examples); (14) compounds with $IrMo_2B_2$ structure type (7 examples); (15) compounds with $BaCu_2S_2$ structure type (7 examples); (16) compounds with this composition and a crystal structure different from those listed above (99 examples), (17) non-formation of compounds with composition AB_2X_2 under ambient conditions (1179 examples).

The attribute set includes 32*3=96 properties of atoms of elements A, B and X (the first three energies of ionization, the thermal conductivities, the molar heat capacities, the electronegativities according to Pauling, the enthalpies of atomization and vaporization, the group and quantum numbers, the numbers of valence electrons, Debye temperatures, Mendeleev numbers, the melting and boiling points, the pseudo-potential radii according to Zunger, the metal radii according to Waber, the covalent radii, the chemical potentials according to Miedema, etc.).

The prediction procedure for the type of the crystal structure of compounds included three tasks:

1. Prediction of A–B–X chemical systems with the formation and non-formation of AB_2X_2 compounds.
2. Multiclass prediction of the type of crystal structure (seventeen above-mentioned classes).
3. For classes 1–3 and 5, successive division of systems into three classes, for example, class 1 - the compounds with $ThCr_2Si_2$ crystal structure type ; class 2 - compounds with the structure different from $ThCr_2Si_2$; and class 3 - the absence of AB_2X_2 compounds in the A–B–X system.

The best algorithms according to the results of examination recognition with cross validation (mostly, these were linear machine methods, k-nearest neighbors, Fisher's linear discriminator, neural networks, and ConFor) were used for collective decision making. The best algorithm for collective decision making when solving this problem and like questions was chosen on the basis of recognition of 100 objects chosen randomly by a uniform distribution. These objects were eliminated from the training sample both during the training process and during the tuning of collective decision making algorithms. By the results of recognition of 100 objects, we evaluated the accuracy of collective decision making. From the set of constructed recognition algorithms and algorithms for collective decision making, we chose a subset of the most accurate algorithms (most frequently, this was the complex committee method—averaging). At the final stage, the pro-

cesses of training of the chosen algorithms were initiated again on the original training sample. Note that the application of collective algorithms has allowed us to substantially increase the reliability of prediction. As a result, for each of 11 AB_2X_2 compositions (A and B are various elements, and X is B, Al, Si, P, Ga, Ge, As, Se, Sn, Sb, or Te), we obtained six tables of predictions (a table of prediction of the possibility of formation of a compound, a table of multiclass prediction, and four tables of predictions for classes 1, 2, and 3 and 5 in which all metal systems are classified into three groups). Next, we compared the predictions from these six tables and made a decision on the existence of a given compound and on the type of its crystal structure provided that the predictions were not contradictory. Table 4 shows a part of predictions obtained in this way for new compounds with composition AB_2Si_2.

Designations: (1) prediction of compound AB_2Si_2 with $ThCr_2Si_2$ crystal structure type; (2) prediction of compound AB_2Si_2 with $CaAl_2Si_2$ crystal structure type; (3) prediction of compound AB_2Si_2 with $CaBe_2Ge_2$ crystal structure type; (4) the crystal structure differing from those listed above is predicted; (5) formation of compound with composition AB_2Si_2 is not predicted; # stands for the designation of objects that have been used in computer training.

Estimation of the Physical Properties of Inorganic Compounds

The prediction of the numerical intrinsic physical properties (for example, melting point of compound at atmospheric pressure, critical temperature of transition to superconducting state, etc.) is the most difficult problem at computer-assisted design of inorganic compounds using computer training. In this case, only a threshold estimation of the property (more or less than a threshold) is possible. The problem is a search for such threshold (and set of attributes) in order to fulfill the basic hypothesis of pattern recognition methods

Table 4. Part of predictions of the crystal structure type for compounds with composition AB_2Si_2

A B	Ca	Sc	Sr	Y	Zr	Ba	La	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu	Th	Pa	U	Np
Mg				2	?	#1	#4	#4																	
Al	#2	2	#2	#2	5		#2	#2	#2	#2	2	#2	#2	#2	#2	#2	#2	#2	2	#2	2	?	?	#4	?
Cr	1	1	1	#1	5	1	1	#5	1	#1	1	#1	1	#1	#1	#1	#1	#1	#1	#1	#1	#1	1	#1	#1
Mn	#1	1	#1	#1		#1	#1	#1	#1	#1	1	#1	1	#1	#1	#1	#1	#1	#1	#1	#1	#1	1	#1	#1
Fe	#5	#4	1	#1	#1	1	#1	#1	#1	#1	1	#1	#1	#1	#1	#1	#1	#1	#1	#1	#1	#1	1	#1	#1
Co	#1	#1	#1	#1	#1	#1	#1	#1	#1	#1	1	#1	#1	#1	#1	#1	#1	#1	#1	#1	#1	#1	1	#1	#1
Ni	#1	#1	#1	#1	#1	1	#1	#1	#1	#1	1	#1	#1	#1	#1	#1	#1	#1	#1	#1	#1	#1	1	#1	#1
Cu	#1	#1	#1	#1	1	#1	#1	#1	#1	#1	1	#1	#1	#1	#1	#1	#1	#1	#1	#1	#1	#1	1	#1	#1
Zn	#1	1	#1	1		#1	1	1	1	#1	1	1	#1	1	1	1	1	1	1	1	1	1	1	1	1
Ru	1	1	1	#1	1		#1	#1	#1	#1	1	#1	1	#1	#1	#1	#1	#1	#1	#1	#1	#1	1	#1	#1
Rh	1	1	1	#1	1	#4	1	#1	1	1	1	1	#1	#1	#1	#1	1	#1	1	#1	#1	#1	1	#1	#1
Pd	#1	1	#1	#1	1		#1	#1	#1	#1	1	#1	#1	#1	#1	#1	#1	#1	#1	#1	#1	#1	1	#1	#1
Ag	#1	1	#1	1		#1	#1	#1	#1	#1	1	#1	#1	#1	1	#1	1	1	1	#1	1	1	1	1	1
Os	1	1	1	1	5	1	1	#1	#1	1	1	1	1	#1	#1	1	#1	#1	1	1	1	#1	1	#1	#1
Ir	1	1	1	#1	1	#4	#1	#1	#1	#1	1	#1	#1	#1	#1	1	1	#1	1	1	1			#3	#3
Pt	#3	1	#3	1		#4	#3	#1	#3	#1		#3	#3	#1		#1	#3	#1	#1	#1	#1			#1	#3
Au	#1	1	#1	#1		#1	#1	#1	#1	#1	1	#1	#1	#1	#1	#1	#1	#1	1	#1	1	#1	1	#1	#3

- hypothesis of compactness. We succeeded in estimating of some physical properties of inorganic compounds (the critical temperature of transition to superconducting state (Savitskii et al., 1979), the melting point and bandgap (Kiselyova et al., 2007a, 2007b), etc.). In particular we predicted wide bandgap semiconductors with chalcopyrite crystal structure using IAS.

The chalcopyrites with composition ABX_2 are promising for the development of new semiconducting, nonlinear optical, and other materials for electronics. The aim of our investigations was to design new semiconducting compounds with crystal structure of chalcopyrite and band gap more than 2 eV for developing optoelectronic devices.

The following parameters of chemical elements were used for the description of chemical compounds:

- The electronegativity in the Martynov-Batsanov scale of values $\Delta\chi$,where $\Delta\chi = \mid 2\chi_C - \chi_A - \chi_B \mid$
- Atomic electrovalent Z_A, Z_B, Z_C (for transition metals, the group number of elements is used as Z);
- Mean Born exponent; $\overline{n} = \frac{n_A + n_B + 2n_C}{4}$;
- The chemical scale χ of Pettifor;
- The proportion: $(I_z / Z)_{AC} = (\frac{I_z}{Z})_A - [6 + 0.1(\frac{I_z}{Z})_C]$, where Iz - final ionization potential;
- Atomic radius.

Data for computer training has been extracted from database "Bandgap" (Kiselyova et al., 2010). The two classes were considered: (1) chalcopyrites with $\Delta E>2$ eV and (2) chalcopyrites with $\Delta E<2$ eV. Table 5 contains experimental values and results of examination of prediction of band gap of known chalcopyrites using separate methods of pattern recognition.

The following methods were used:

- **EC**: Estimate calculation algorithms,
- **LDF**: Fisher's linear discriminant,
- **LM**: Linear machine,
- **MR**: Algorithm of logical regularities,
- **NN**: Neural networks,
- **KNN***: k*–nearest neighbors,
- **SVM**: Support vector machine,
- **SWS**: Statistically weighted syndromes,
- **TA**: Deadlock test algorithm,
- **LG**: Logical regularities of recognized object,
- **DT**: Method of binary decisive trees,
- **CF**: ConFor,
- **SVR**: Support vector regression.

Recognition uses procedure of cross-validation (excepting CF and SVR). The best results were achieved using algorithms of logical regularities, linear machine and ConFor (in the last case the examination recognition of 100 objects chosen randomly by a uniform distribution and eliminated from the training sample was used for determination of predicting accuracy). The prediction of band gap of new chalcopyrites was carried out using results of algorithms of logical regularities and linear machine on the basis of application of the following collective methods to making a decision:

- **BM**: Bayesian method,
- **C&S**; Clustering and selection,
- **DT**: Decision templates,
- **WDM:** Woods dynamic method,
- **CS**: Convex stabilizer,
- **CM-MV**: Committee method - majority voting,
- **CM-A**: Committee method – average value,
- **LC**: Logical correction.

Previous control recognition showed that the best results of collective recognition could be

Table 5. Prediction of ΔE of known chalcopyrites (threshold = 2 eV)

Compound	Experimental		Prediction												
	class	ΔE, eV	EC	LDF	LM	MR	NN	KNN	SVM	SWS	TA	LG	DT	CF	SVR
$CuAlS_2$	1	3.5	1	1	1	1	1	1	1	1	1	1	1	1	1
$CuGaS_2$	1	2.44	2	1	1	1	1	1	1	2	2	1	2	1	1
$CuInS_2$	2	1.5	2	2	1	?	2	2	2	2	2	1	2	2	2
$CuAlSe_2$	1	2.67	2	1	1	1	1	1	1	2	2	1	2	1	1
$CuGaSe_2$	2	1.63	2	1	1	1	1	1	1	2	2	1	2	2	2
$CuInSe_2$	2	0.95	2	2	2	2	2	2	2	2	2	2	2	2	2
$CuAlTe_2$	1	2.06	2	2	2	2	1	1	1	2	2	2	2	1	1
$CuGaTe_2$	2	1.18	2	2	2	2	1	1	2	2	2	2	2	2	2
$CuInTe_2$	2	0.88	2	2	2	2	2	2	2	2	2	2	2	2	2
$AgAlS_2$	1	3.13	2	1	1	1	1	1	1	1	1	1	1	1	1
$AgGaS_2$	1	2.75	2	1	1	1	1	1	1	2	1	1	2	1	1
$AgAlSe_2$	1	2.55	2	1	1	1	1	1	1	2	1	1	2	1	1
$AgGaSe_2$	2	1.65	2	1	2	2	1	1	1	2	2	2	2	2	2
$AgInSe_2$	2	1.24	2	2	2	2	2	2	2	2	2	2	2	2	2
$AgAlTe_2$	2	1.8	2	2	2	2	1	1	2	2	2	2	2	2	2
$AgGaTe_2$	2	1.1	2	2	2	2	1	2	2	2	2	2	2	2	2
$AgInTe_2$	2	0.96	2	2	2	2	2	2	2	2	?	2	2	2	2
$ZnSiP_2$	1	2.07	2	2	1	1	1	2	2	1	1	1	2	1	1
$ZnSiAs_2$	1	2.1	2	2	1	1	1	2	2	1	1	1	2	1	1
$ZnGeN_2$	1	2.9	2	2	1	1	1	1	2	1	1	1	2	1	1
$ZnGeP_2$	1	2.1	2	2	1	1	1	2	2	1	1	1	2	1	1
$ZnGeAs_2$	2	1.16	2	2	2	2	2	2	2	2	2	2	2	2	2
$ZnSnP_2$	2	1.45	2	2	2	2	2	2	2	2	2	2	2	2	2
$ZnSnAs_2$	2	0.74	2	2	2	2	2	2	2	2	2	2	2	2	2
$ZnSnSb_2$	2	0.4	2	2	2	2	2	2	2	2	2	2	2	2	2
$CdSiP_2$	1	2.2	2	2	1	1	1	2	2	1	1	1	2	1	1
$CdGeP_2$	2	1.8	2	2	2	2	1	2	2	1	1	2	2	2	2
$CdGeAs_2$	2	0.53	2	2	2	2	1	2	2	2	2	2	2	2	2
$CdSnP_2$	2	1.16	2	2	2	2	2	2	2	2	2	2	2	2	2
$CdSnAs_2$	2	0.3	2	2	2	2	2	2	2	2	2	2	2	2	2
$AgInS_2$	2	1.9	2	2	2	2	2	2	2	2	2	2	2	2	2
$CdSiAs_2$	2	1.51	2	2	2	2	1	2	2	1	1	2	2	2	2
$CuFeS_2$	2	0.53	1	2	2	1	1	1	2	1	1	1	2	2	2
$CuFeSe_2$	2	0.16	2	2	2	?	1	2	2	2	2	2	2	2	2
$CuFeTe_2$	2	0.1	2	2	2	2	1	2	2	2	2	2	2	2	2
$LiGaTe_2$	1	2.31	2	2	2	1	2	2	2	2	2	2	2	1	1
$LiInTe_2$	2	1.46	2	2	2	2	2	2	2	2	2	2	2	2	2
$AgFeSe_2$	2	0.23	2	2	2	2	1	2	2	2	2	2	2	2	2
$MgSiP_2$	1	2.35	2	1	1	1	1	2	2	1	?	1	2	1	1
$MnGeP_2$	2	0.24	2	2	2	1	1	2	2	1	1	1	2	2	2
$MnGeAs_2$	2	0.6	2	2	2	2	2	2	2	2	2	2	2	2	2

Table 6. Prediction of ΔE of new chalcopyrites (calculations using collective methods) (threshold = 2 eV)

Compound	Prediction of ΔE	BM	C&S	DT	WDM	CS	CM-MV	CM-A	LC
$ZnAlS_2$	1	1	1	1	?	1	1	1	1
$ZnAlSe_2$	1	1	1	1	1	1	1	1	1
$ZnAlTe_2$	2	2	2	2	?	2	2	2	2
$AgFeS_2$	2	2	2	2	2	2	2	2	2
$AgFeTe_2$	2	2	2	2	2	2	2	2	2
$ZnGaTe_2$	2	2	2	2	?	2	2	2	2
$CdGaTe_2$	2	2	2	2	?	2	2	2	2
$HgGaTe_2$	2	2	2	2	?	2	2	2	2
$BeCN_2$	1	1	1	1	1	1	1	1	1

achieved using Bayesian method and convex stabilizer strategies (error of predicting equals 0%). The results of these algorithms were used for making a decision at prediction of band gap of new chalcopyrites (Table 6). Thus three new compounds ($ZnAlS_2$, $ZnAlSe_2$ and $BeCN_2$) are promising for opto-electronic applications (Table 6). The following designations were used: (1) chalcopyrites with $\Delta E>2$ eV and (2) chalcopyrites with $\Delta E<2$ eV.

CONCLUSION

The application of machine training and pattern recognition methods to the computer-assisted design of inorganic compounds allows one to find complex classification regularities that make it possible to predict the membership of new chemical systems in one class of substances or another on the basis of knowledge of the well-known properties of the components of these systems — chemical elements. Using these methods it was possible to carry out the prediction of thousands of new compounds and estimation of some of their properties. Computer-assisted design allows one to substantially reduce the number of complex and expensive experiments in the search for inorganic compounds with predefined properties, replacing them by computation. The experimental verification of the results of computer-assisted design shows that the average accuracy of predicting is higher than 80%.

ACKNOWLEDGMENT

This work was partially supported by the Russian Foundation for Basic Research (project nos. 06-07-89120, 08-01-90427, 08-07-00437, 05-03-39009, 11-07-00715, and 09-07-00194). We are grateful to V. P. Gladun and V. Yu. Velichko for long-term help and collaboration.

REFERENCES

Aleksovska, S., Dimitrovska, S., & Kuzmanovski, I. (2007). Crystal structure prediction in orthorhombic ABO_3 perovskites by multiple linear regression and artificial neural networks. *Acta Chimica Slovenica*, *54*(3), 574–582.

Bahrami, A., Mousavi, A. S. H., & Ekrami, A. (2005). Prediction of mechanical properties of DP steels using neural network model. *Journal of Alloys and Compounds*, *392*(1-2), 177–182.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. New York, NY: Chapman and Hall.

Burkhanov, G. S., & Kiselyova, N. N. (2009). Prediction of intermetallic compounds. *Russian Chemical Reviews*, *78*(6), 569–587.

Cai, K., Xia, J., Li, L., & Gui, Z. (2006). Analysis of the electrical properties of PZT by a BP artificial neural network. *Computational Materials Science*, *34*(2), 166–172.

Ceder, G., Morgan, D., Fischer, C., Tibbetts, K., & Curtarolo, S. (2006). Data-mining-driven quantum mechanics for the prediction of structure. *MRS Bulletin*, *31*(8), 981–985.

Chen, N., Chen, R., Lu, W., Li, C., & Villars, P. (1999). Regularities of formation of ternary intermetallic compounds. Part 4. Ternary compounds between two nontransition elements and one transition element. *Journal of Alloys and Compounds*, *292*(1-2), 129–133.

Chen, N., Li, C., & Qin, P. (1998). KDPAG expert system applied to materials design and manufacture. *Engineering Applications of Artificial Intelligence*, *11*(5), 669–674.

Chen, N., Li, C., Yao, S., & Wang, X. (1996). Regularities of melting behavior of some binary alloy phases. Part 1. Criteria for congruent and incongruent melting. *Journal of Alloys and Compounds*, *234*(1-2), 125–129.

Chen, N., Li, C., Yao, S., & Wang, X. (1996). Regularities of melting behavior of some binary alloy phases. Part 2. Computerized prediction of melting points of alloy phases. *Journal of Alloys and Compounds*, *234*(1-2), 130–136.

Chen, N., Lu, W., Chen, R., Qin, P., & Villars, P. (1999). Regularities of formation of ternary intermetallic compounds. Part 1. Ternary intermetallic compounds between nontransition elements. *Journal of Alloys and Compounds*, *289*(1-2), 120–125.

Chen, N., Lu, W., Qin, P., Chen, R., & Villars, P. (1999). Regularities of formation of ternary intermetallic compounds. Part 2. Ternary compounds between transition elements. *Journal of Alloys and Compounds*, *289*(1-2), 126–130.

Chen, N., Zhu, D. D., & Wang, W. (2000). Intelligent materials processing by hyperspace data mining. *Engineering Applications of Artificial Intelligence*, *13*(5), 527–532.

Chen, N. Y., Lu, W. C., Yang, J., & Li, G. Z. (2004). *Support vector machine in chemistry*. Singapore: World Scientific Publishing Co. Pte. Ltd.

Devingtal, Y. V. (1968). About optimal coding of objects at their classification using pattern recognition methods. *Proceedings of Academy of Sciences of USSR, Technical Cybernetics (Izvestiya Akademii Nauk SSSR. Tekhnicheskaya Kibernetika),* 1, (pp. 162-169). (in Russian)

Devingtal, Y. V. (1971). Coding of objects at application of separating hyper-plane for their classification. *Proceedings of Academy of Sciences of USSR. Technical Cybernetics (Izvestiya Akademii Nauk SSSR. Tekhnicheskaya Kibernetika),* 3, (pp. 139-147). (in Russian)

Gladun, V. P. (1972). Concept formation by learning in growing networks. *Cybernetics and Systems Analysis*, *6*(2), 124–130.

Gladun, V. P. (1995). *Processes of formation of new knowledge.* Sofia, Bulgaria: SD "Pedagog 6" (in Russian)

Gladun, V. P. (1997). Hypothetical modeling: Methodology and application. *Cybernetics and Systems Analysis*, *33*(1), 7–15.

Gladun, V. P., & Vashchenko, N. D. (1975). Methods for forming concepts with a computer [review]. *Cybernetics and Systems Analysis*, *11*(2), 295–301.

Gribov, L. A. (2010). Some conceptual issues for the statement of quantum problems in the theory of molecular structure and transformations. *Journal of Structural Chemistry*, *51*(4), 603–615.

Gu, T., Lu, W., Bao, X., & Chen, N. (2006). Using support vector regression for the prediction of the band gap and melting point of binary and ternary compound semiconductors. *Solid State Sciences*, *8*(2), 129–136.

Gulyaev, B. B., & Pavlenko, L. F. (1973). Modelling of the search for alloy dopants. *Automation and Telemechanics*, *1*, 131–134.

Hautier, G., Fischer, C. C., Jain, A., Mueller, T., & Ceder, G. (2010). Finding nature's missing ternary oxide compounds using machine learning and density functional theory. *Chemistry of Materials*, *22*(12), 3762–3767.

Hayajneh, M., Hassan, A. M., Alrashdan, A., & Mayyas, A. T. (2009). Prediction of tribological behavior of aluminum–copper based composite using artificial neural network. *Journal of Alloys and Compounds*, *470*(1-2), 584–588.

Iserman, M. A., Braverman, E. M., & Rosonoer, L. I. (1970). *The method of potential functions in the theory of machine training.* Moscow, Russsia: "Nauka" (in Russian)

Kiselyova, N., Stolyarenko, A., Ryazanov, V., & Podbel'skii, V. (2008). Information-analytical system for design of new inorganic compounds. *International Journal ". Information Theories & Applications*, *2*(4), 345–350.

Kiselyova, N. N. (1993a). Prediction of inorganic compounds: Experiences and perspectives. *MRS Bulletin*, *18*(2), 40–43.

Kiselyova, N. N. (1993b). Information-predicting system for the design of new materials. *Journal of Alloys and Compounds*, *197*(2), 159–165.

Kiselyova, N. N. (2000). Databases and semantic networks for the inorganic materials computer design. *Engineering Applications of Artificial Intelligence*, *13*(5), 533–542.

Kiselyova, N. N. (2002). Computer design of materials with artificial intelligence methods. In J. H. Westbrook & R. L. Fleischer (Ed.), *Intermetallic compounds. Principles and practice. Vol.3. Progress* (pp. 811-839). Chichester, UK: John Wiley&Sons, Ltd.

Kiselyova, N. N. (2005). *Computer design of inorganic compounds. Application of databases and artificial intelligence*. Moscow, Russia: Nauka. (in Russian)

Kiselyova, N. N., Dudarev, V. A., & Zemskov, V. S. (2010). Computer information resources in inorganic chemistry and materials science. *Russian Chemical Reviews*, *79*(2), 145–166.

Kiselyova, N. N., Gladun, V. P., & Vashchenko, N. D. (1998). Computational materials design using artificial intelligence methods. *Journal of Alloys and Compounds*, *279*(1), 8–13.

Kiselyova, N. N., LeClair, S. R., Gladun, V. P., & Vashchenko, N. D. (2000). Application of pyramidal networks to the search for new electro-optical inorganic materials. In I. J. Rudas & J. K. Tar (Ed.), *IFAC Symposium on Artificial Intelligence in Real Time Control AIRTC-2000, Preprints,* Budapest, Hungary, October 2-4, 2000 (pp. 35-40). Budapest, Hungary: Budapest Polytechnic.

Kiselyova, N. N., Pokrovskii, B. I., Komissarova, L. N., & Vashchenko, N. D. (1977). Modelling of formation of complicated oxides from the initial components on the basis of cybernetic method of concept formation. *Zhurnal Neorganicheskoi Khimii, (English translation - Russian Journal of Inorganic Chemistry), 22*(4), 883-886 (in Russian)

Kiselyova, N. N., Prokoshev, I. V., Dudarev, V. A., Khorbenko, V. V., Belokurova, I. N., Podbel'skii, V. V., & Zemskov, V. S. (2004). Internet-accessible electronic materials database system. *Inorganic Materials*, *40*(3), 321–325.

Kiselyova, N. N., Ryazanov, V. V., & Sen'ko, O. V. (2009). Prediction of the types of crystal structure for ABX_2 (X = Fe, Co, Ni) intermetallics. *Russian Metallurgy (Metally)*, *6*, 538–545.

Kiselyova, N. N., & Savitskii, E. M. (1983). Prediction of formation of ternary silicides with $ThCr_2Si_2$ structure type. *Izvestiya Akademii Nauk SSSR. Neorganicheskie Materialy*, *19*(3), 489–491.

Kiselyova, N. N., Stolyarenko, A. V., Gu, T., & Lu, W. (2007a). Computer-aided design of new wide bandgap semiconductors with chalcopyrite structure. *Advanced Materials (Deerfield Beach, Fla.)*, (Special Issue), 351–355.

Kiselyova, N. N., Stolyarenko, A. V., Gu, T., Lu, W., Blansche, A., Ryazanov, V. V., & Senko, O. V. (2007b). Computer-aided design of new inorganic compounds promising for search for electronic materials. In *Proceedings of the Sixth International Conference on Computer-Aided Design of Discrete Devices(CAD DD 07). Vol.1.* (pp. 236-242). Minsk, Belarus: UIPI NASB.

Kiselyova, N. N., Stolyarenko, A. V., Ryazanov, V. V., Senko, O. V., Dokukin, A. A., & Podbel'skii, V. V. (2011). A system for computer-assisted design of inorganic compounds based on computer training. *Pattern Recognition and Image Analysis*, *21*(1), 88–94.

Kockan, U., & Evis, Z. (2010). Prediction of hexagonal lattice parameters of various apatites by artificial neural networks. *Journal of Applied Crystallography*, *43*(4), 769–779.

Kohanoff, J. (2006). *Electronic structure calculations for solids and molecules: Theory and computational methods*. Cambridge, UK: Cambridge University Press.

Kovshov, N. V., Moiseev, V. L., & Ryazanov, V. V. (2008). Algorithms for finding logical regularities in pattern recognition. *Computational Mathematics and Mathematical Physics*, *48*(2), 314–328.

Kutolin, S. A., Komarova, S. N., & Frolov, Y. A. (1982). Prediction of the type of defects and their concentration and energetics of defect formation of imperfect crystals of refractory compounds as function of electronic structure and composition using computer. *Zhurnal fizicheskoi khimii (English translation - Russian Journal of Physical Chemistry A, Focus on Chemistry), 56*(4), 996-999. (in Russian)

Kutolin, S. A., & Kotyukov, V. I. (1978). Function of the chemical affinity and prediction of binary compositions and properties of rare earth compounds using computer. *Zhurnal Fizicheskoi Khimii (English translation - Russian Journal of Physical Chemistry A, Focus on Chemistry), 52*(4), 918-922 (in Russian)

Kutolin, S. A., & Kotyukov, V. I. (1979). Prediction of composition of compounds in ternary systems and of their properties as function of electronic structure of components using computer. *Izvestiya Akademii Nauk SSSR* [Inorganic Materials]. *Neorganicheskie Materialy*, *15*(8), 1389–1401.

Kutolin, S. A., Vashukov, I. A., & Kotyukov, V. I. (1978). Prediction of binary rare earth compounds and of their properties using computer. *Izvestiya Akademii Nauk SSSR. Neorganicheskie Materialy (English translation - Inorganic Materials), 14*(2), 215-218. (in Russian)

Lenovich, A. S. (1974). An algorithm of recognition of the effect of the chemical composition of steel on its mechanical properties. *Cybernetics and Systems Analysis*, *10*(3), 542–544.

Li, Y. (2006). Predicting materials properties and behavior using classification and regression trees. *Materials Science and Engineering A*, *433*(1-2), 261–268.

Liu, H.-L., Chen, Y., & Chen, N.-Y. (1994). PCB method applied to material design - computer aided synthesis of BiPbSrCaCuOF superconductor. *Journal of Chemometrics*, *8*(5), 439–443.

Lu, W., Chen, N., Li, C., Qin, P., Chen, R., Yao, L., & Tao, L. (1999). Regularities of formation of ternary intermetallic compounds. Part 3. Ternary intermetallic compounds between one nontransition element and two transition elements. *Journal of Alloys and Compounds*, *289*(1-2), 131–134.

Ryazanov, V. V. (2007). Logical regularities in pattern recognition (parametric approach). *Computational Mathematics and Mathematical Physics*, *47*(10), 1720–1735.

Savitskii, E. M., Devingtal, Y. V., & Gribulya, V. B. (1968a). About recognition of binary state diagrams of metallic systems using computer. *Dokl. Akad. Nauk SSSR (English translation - Doklady Physical Chemistry), 178*(1), 79-81. (in Russian)

Savitskii, E. M., Devingtal, Y. V., & Gribulya, V. B. (1968b). Prediction of metallic compounds with composition A_3B using computer. *Dokl. Akad. Nauk SSSR (English translation - Doklady Physical Chemistry), 183*(5), 1110-1112 (in Russian)

Savitskii, E. M., & Gribulya, V. B. (1985). *Application of computer techniques in the prediction of inorganic compounds*. New Delhi, India: Oxonian Press Pvt., Ltd.

Savitskii, E. M., Gribulya, V. B., & Kiseleva, N. N. (1981). Forecasting of superconducting compounds. [a]. *Physica Status Solidi*, *63*(1), K67–K72.

Savitskii, E. M., Gribulya, V. B., & Kiselyova, N. N. (1979). Cybernetic prediction of superconducting compounds. *Calphad*, *3*(3), 171–173.

Savitskii, E. M., Gribulya, V. B., & Kiselyova, N. N. (1980). On the application of cybernetic prediction systems in the search for new magnetic materials. *Journal of the Less Common Metals*, *72*(2), 307–315.

Savitskii, E. M., Gribulya, V. B., Kiselyova, N. N., Ristic, M., Nikolic, Z., & Stoilkovic, Z. (1990). *Prediction in materials science using computer*. Moscow, Russia: Nauka. (in Russian)

Savitskiy, E. M., Gribulya, V. B., & Kiselyova, N. N. (1982). Cybernetic prediction of inorganic compounds and its correlation with experiment. *Crystal Research and Technology*, *17*(1), 3–17.

Scott, D. J., Manos, S., & Coveney, P. V. (2008). Design of electroceramic materials using artificial neural networks and multiobjective evolutionary algorithms. *Journal of Chemical Information and Modeling*, *48*(2), 262–273.

Senko, O. V. (2009). An optimal ensemble of predictors in convex correcting procedures. *Pattern Recognition and Image Analysis*, *19*(3), 465–468.

Senko, O. V., & Dokukin, A. A. (2010). Optimal forecasting based on convex correcting procedures. In *New trends in classification and data mining* (pp. 62–72). Sofia, Bulgaria: ITHEA.

Talanov, V. M., & Frolova, L. A. (1979). The investigation of possibility of formation of chemical compounds with spinel structure using method of potential functions. *Reports of Universities – Chemistry and Chemical Technology (Izvestiya VUZov. Khimia i Khimicheskaya Tekhnologiya), 22*(9), 1044-1047. (in Russian)

Talanov, V. M., & Frolova, L. A. (1981). The investigation of possibility of formation of chalcospinel structure using method of potential functions. *Reports of Universities – Chemistry and Chemical Technology (Izvestiya VUZov. Khimia i Khimicheskaya Tekhnologiya), 24*(3), 274-276 (in Russian)

Taskin, M., Dikbas, H., & Caligulu, U. (2008). Artificial neural network (ANN) approach to prediction of diffusion bonding behavior (shear strength) of Ni-Ti alloys manufactured by powder metallurgy method. *Mathematical and Computational Applications*, *13*(3), 183–191.

Thaler, S. L. (1998). Predicting ultra-hard binary compounds via cascades auto- and hetero-associative neural networks. *Journal of Alloys and Compounds*, *279*(2), 47–59.

Vozdvizhenskii, V. M. (1974). *Prediction of binary state diagrams*. Moscow, Russia: Metallurgiya. (in Russian)

Vozdvizhenskii, V. M., & Falevich, V. Y. (1973). Application of pattern recognition method to determination of type of phase diagram of binary metallic systems. In *General behaviour in the structure of state diagrams of metallic systems* (pp. 119–120). Moscow, Russia: Nauka. (in Russian)

Yan, L.-M., Zhan, Q.-B., Qin, P., & Chen, N.-Y. (1994). Study of properties of intermetallic compounds of rare earth metals by artificial neural networks. *Journal of Rare Earths*, *12*(2), 102–107.

Yao, L., Qin, P., Chen, N., & Villars, P. (•••). (200l). TICP -An expert system applied to predict the formation of ternary intermetallic compounds. *Calphad*, *25*(1), 27–30.

Ye, H., Xia, B., Wu, W., Du, K., & Zhang, H. (2002). Effect of rare earth composition on the high-rate capability and low-temperature capacity of AB_5-type hydrogen storage alloys. *Journal of Power Sources*, *111*(1), 145–151.

Zeng, Y., Chua, S. J., & Wu, P. (2002). On the prediction of ternary semiconductor properties by artificial intelligence methods. *Chemistry of Materials*, *14*(7), 2989–2998.

Zhou, B., Jin, S. M., Shao, J., & Chen, N. (1989). IMEC – An expert system for retrieval and prediction of binary intermetallic compounds. *Acta Metallurgica Sinicae Serie B*, *2*(6), 428–433.

Zhuravlev, Y. I., Kiselyova, N. N., Ryazanov, V. V., Senko, O. V., & Dokukin, A. A. (2011). Design of inorganic compounds with the use of precedent-based pattern recognition methods. *Pattern Recognition and Image Analysis*, *21*(1), 95–103.

Zhuravlev, Y. I., & Nikiforov, V. V. (1971). Recognition algorithms based on computation of estimates. *Cybernetics and Systems Analysis*, *7*(3), 387–400.

Zhuravlev, Y. I., Ryazanov, V. V., & Sen'ko, O. V. (2006). *RECOGNITION: Mathematical methods, software system, practical solutions*. Moscow, Russia: Phasis. (in Russian)

KEY TERMS AND DEFINITIONS

Concept: A generalized model of some class of objects that provides for recognizing and generating models of specific elements of this class.

CONFOR: (CONcept FORmation): A set of software tools intended for the logical analysis of large volumes of experimental data (Gladun, 1997, 1995, 1972; Gladun & Vashchenko, 1975) using the special computer memory structure – growing network - with the purpose of searching for regularities.

Design of New Inorganic Compounds: A search for combination of chemical elements and their ratio (i.e., determining qualitative and quantitative compositions) for the synthesis (under given conditions) of the predefined space molecular or crystal structure of a compound that possesses the required functional properties. It is the knowledge of the properties of chemical elements and data about other compounds already investigated that constitute initial information for the calculations.

Information-Analytical System (IAS): A system intended for data retrieval on known compounds, predicting inorganic compounds not yet synthesized, and the forecasting of their properties. This system employs databases on properties of inorganic compounds and materials, a database of elements' properties, a subsystem of data analysis based on algorithms of pattern recognition, a subsystem of selecting the most important attributes, a visualization subsystem, a knowledge base (tasks base), a predictions base, and a managing subsystem (Figure 1).

Inorganic Compound: A compound which does not contain carbon (except for carbides, cyanides, carbonates, carbon oxides, and some other compounds that attributed traditionally to inorganic substances).

Prediction: An identification (classification) of a new object belonging to a certain class in compliance with a fixed classification scheme.

Quality (Accuracy) of the Algorithm: A percentage of correctly recognized objects.

RECOGNITION: A set of software tools developed at A. A. Dorodnicyn Computing Centre, Russian Academy of Sciences (CC RAS) (Zhuravlev et al., 2006). This multifunctional system of pattern recognition includes the well-known methods of *k*–nearest neighbors, Fisher's linear discriminator, linear machine, neural networks, support vector machine, genetic algorithm, and the special algorithms developed by the CC RAS: estimate calculating algorithms, LoReg (Logical Regularities), deadlock test algorithm, statistical weighted syndromes, etc. The system contains also a set of collective methods for final decision making (algebraic, logical, and heuristic correctors) and software for cluster-analysis.