Journal of Alloys and Compounds, 197 (1993) 159–165
JALCOM 7317

Information-predicting systems for the design of new materials

N.N. Kiselyova
A.A. Baikov Institute of Metallurgy, Russian Academy of Sciences, Leninskiy Prospect 49, 117334 Moscow (Russian Federation)

Abstract

In this paper we discuss the main principles for the cybernetic prediction of inorganic substances which would have predefined properties. These predictive techniques are based on machine learning strategies. The efficiency of the proposed approach is illustrated by comparing the results of predicting the properties of new substances with experimental data. Examples showing the application of these techniques to the prediction of those materials used in the electronics industry are described. Also discussed are the components and the organization of the computer-aided information-predicting system developed by the Institute of Metallurgy of the Russian Academy of Sciences. A detailed description of the databank for the properties of the ternary inorganic phases is provided, as well as an explanation of the main difference between the information-predicting system and the expert systems.

1. Introduction

The problem of the computer-aided design of new materials has two aspects. The first is the prediction of compounds, not yet synthesized, which would have predefined properties. The second aspect is the search for optimum conditions for the production of the new materials with the aim of optimization of the desired properties. In most cases these two problems can be successfully solved by applying the cybernetic–statistical approach as outlined by Savitskii *et al.* [1]. The principal idea of this approach is the combination of the cybernetic prediction methods used to solve the first problem and the statistical methods of design of the required experiments (or other methods of optimizing) that provide the solution for the second problem. Experiments aimed at the application of such an approach have been successful and described elsewhere [1].

In this study we describe the cybernetic prediction of substances (with desired properties), not yet synthesized, and the extension of this approach to designing a new class of information-predicting systems, for use in materials science.

Initially the idea of machine learning, to recognize patterns, for predicting new inorganic binary phases was proposed by Savitskii *et al.* [2]. These concepts of machine learning emerged as a result of the advances, at that time, in computational power, and in the application of this computational ability to process information. These developments aided the implementation of algorithms for the prediction of unknown inorganic phases. This method of predicting new materials using machine learning strategies has been termed the "cybernetic prediction of inorganic compounds" [1], to distinguish it from quantum mechanical, statistical and other empirical approaches [3–7].

2. Concepts for the cybernetic approach

The foundations of the cybernetic prediction of inorganic compounds are based on Mendeleev's law which asserts that the periodic nature of changes in the properties of chemical systems depends on the nature and properties of the elements which make these systems (compounds, solutions and the like). Another premise justifying the proposed approach is the existence of good classification schemes for inorganic substances. The problem of predicting new compounds was reduced to the analysis of the multidimensional array of property values and the column vector of the desired properties. Each row corresponds to some known physicochemical system, whose class is indicated by the row position of the column vector. The process of analysing all this information, aimed at finding the classifying regularities, with the help of special programs is referred to as machine learning. By substituting the values of the properties of the components of the unknown system in the regularity thus found, it is possible to determine the class. The implementation of this stage, called the prediction, requires only the knowledge of the values of the component properties.

The problem of prediction is easy to solve if there are enough objects (examples of systems belonging to

different classes) for machine learning, and the classification is of a qualitative nature (*e.g.* type of phase diagram, crystal chemical classification and the like). The problem is more difficult to solve when some quantitative values are to be estimated (for example, the phase transition temperature and pressure formation enthalpy). Before applying machine learning strategies in such a case it is advisable to find the threshold values of the object function characteristic of the particular data array. In the latter case the problem is, as a rule, solved by the methods of cluster analysis (automatic classification). However, sometimes, as in the case of estimating the critical temperature of transition to the superconducting state, for example, it becomes necessary to proceed from the practical requirements and to use the known threshold values, *e.g.* boiling point temperatures of helium and nitrogen.

To discuss, in brief, the methods of machine learning, it must be pointed out that since there are very many aspects in this domain of artificial intelligence no criteria for selecting the most suitable algorithm for a particular application are available. After testing many programs intended for machine learning applications we fixed on the class of algorithms in which all classifying regularities which are found could be presented in the form of a boolean expression. This choice proved to be correct, since it is acceptable practice when developing expert systems to make use of logical functions to represent the knowledge.

The implementation of the machine learning logical algorithms directly involves decreasing the number of searches in the process of forming the classifying regularities. The algorithm's inductive concept formation [8], which we have used for nearly 20 years, solves this particular problem by structuring the computer memory in the form of semantically growing pyramidal networks. The construction of such a network during the input of the objects is followed by studying it, *i.e.* by singling out the control nodes and determining the class to which the objects belong. It must be remembered that each object (physicochemical system) is put in as a set of values of the component properties with an indication of the class to which the system belongs. The resultant classifying regularities can be stored in computer memory and printed or read out in the form of a learned semantic network or an equivalent boolean expression in which the values of the component properties make the variables. During the prediction process the computer receives the atomic numbers of the elements or designations of simple compounds, while the values of the properties of the appropriate elements or simple compounds are automatically extracted from the databank. After their sampling by means of a special program [9] they are substituted in the classifying regularity in which they fall and the researcher can easily obtain the necessary prediction.

An important problem of any computer classification is the choice of the properties for the description of physicochemical systems. This procedure can hardly be formalized completely, but the machine learning algorithm we use automatically rejects those properties which have no importance for the classification process. The initial set of the properties for the computer-aided analysis is prepared by the materials scientist and it is desirable that the artificial intelligence system extrapolates information from this representative set of characteristics.

One interesting feature of these algorithms [8, 9], specifically adapted for chemical application, is their ability to function adequately with incomplete property information or ranges where no information exists, *e.g.* melting point of carbon at atmospheric pressure.

3. Choice of property

We have shown [10] that all fundamental properties of the chemical elements are essential for this machine learning process, *e.g.* distribution of electrons over the energy shells of isolated atoms, ionization potentials, ionic, covalent and atomic radii, melting points, and standard entropies of individual substances. In addition, all properties of simple compounds, oxides, chalcogenides, halides, etc., are included as required by the composition of the compounds to be predicted.

Using these concepts we have successfully solved the following problems [10–15]: (1) prediction of compound formation or non-formation for ternary systems; (2) prediction of the possibility of forming ternary and more complicated compounds of desired composition; (3) prediction of phases with definite crystal structures; (4) estimation of phase quantitative properties (critical temperature of transition to superconducting state, homogeneity region etc.).

Unless otherwise specified, the prediction is conducted for the physicochemical systems at normal conditions, for example, the prediction of a phase at normal pressure and room temperature. In order to predict compounds which exist under different experimental conditions (*e.g.* high pressure), it is necessary to enter the data on known compounds which exist under those experimental conditions (*e.g.* high pressure) into the "learning step" and to add the experimentals condition (*e.g.* pressure) as a parameter into the description. Unfortunately, a majority of the experimental measurements of compound properties are conducted for incomplete equilibrium. In addition, the determination of the crystal structure of a given compound is not, very often, undertaken in conjunction with its phase

diagram investigation. Therefore, it is not always clear under what conditions a specific crystal modification of a compound (polymorphism) is stable.

By applying the cybernetic approach described above we have predicted the formation of thousands of new compounds in ternary, quaternary and more complicated systems. These compounds were then searched for new semiconductors, superconductors, ferroelectrics, magnets and other materials required for new technologies [1, 10–15]. The comparison of these predictions with the experimental data, obtained later, showed [10] that the average reliability of predicted ternary compounds exceeds 90%. Such a high accuracy for *a priori* predictions is not ensured by any other known theoretical method.

Illustrated in Fig. 1 is the comparison between the results of predicting the compounds of the form ABS_2 [11] and the new experimental data. These compounds are of some interest in the search for new semiconductor materials. Out of 29 predictions checked, only 2 were detected to be in error. Recently we have reconfirmed the results of the predictions of the phases of this composition using new data [10]. In the process of our further studies aimed at searching for new magnetic substances [12] we predicted compounds with their crystal structure resembling the Heusler alloys and $ThCr_2Si_2$. Shown in Fig. 2 is a part of the table illustrating predictions of the Heusler alloys with composition AB_2In [13]. Of the 21 compounds studied recently, 19 compounds coincided with those predicted by us. Figure 3 shows the results of comparing the predicted phases of composition AB_2Si_2 with structure type $ThCr_2Si_2$ [14] with the experimental data. Of 71 checked predictions, 68 agreed with the experimental data, *i.e.* a prediction error of 4%.

The results of predicting new superconductors [1, 10, 15] are also significant practically. Recently we have attempted to predict the compounds with the structure of T, T′ and T* phases belonging to promising classes of high temperature superconductors.

The best criterion of correctness for any theoretical method is the experimental validation of the prediction. The experimental validation of our theoretical results has shown that this method of cybernetic prediction is a useful tool for materials scientists and chemists concerned with the *a priori* prediction of inorganic substances which would have predefined properties.

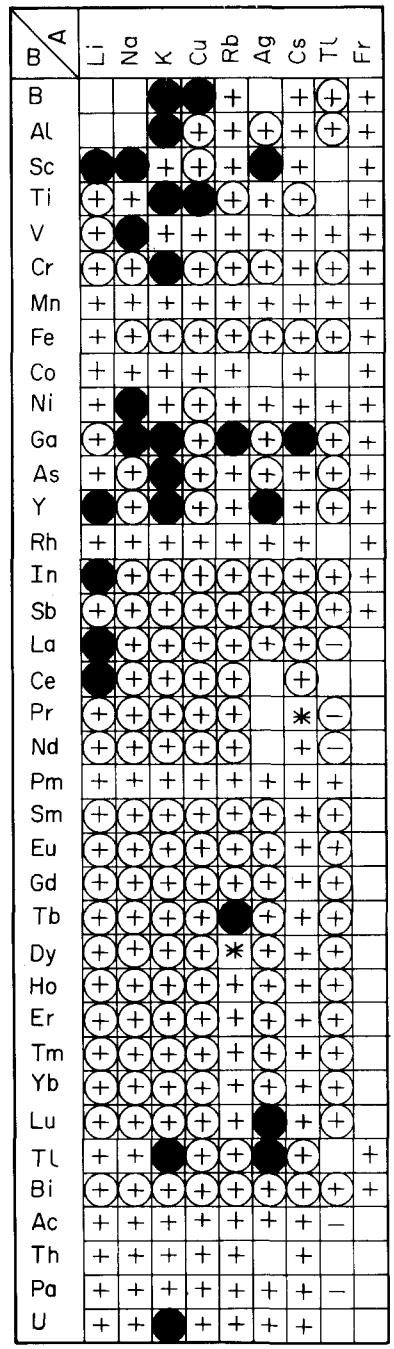

Fig. 1. Part of a two-dimensional plot illustrating the prediction of compounds with the composition ABS_2: +, predicted formation of compound with composition ABS_2; −, prediction of no compound with composition ABS_2; ⊕, compound ABS_2 is formed and this fact is used in the machine learning process; ⊖, compound ABS_2 is not formed and this fact is used in the machine learning process; ●, predicted formation of compound with composition ABS_2 which is confirmed by experiment; ⋆, predicted formation of compound with composition ABS_2 which is not confirmed by experiment.

4. Information-predicting computer system

For the purpose of our studies in this field we developed an information-predicting computer system. This system is based on cybernetic predicting concepts, discussed in detail above, and employs a database of inorganic compounds properties and a predicting subsystem (see Fig. 4). The system is developed for the IBM PS/2 computer operating under MS-DOS.

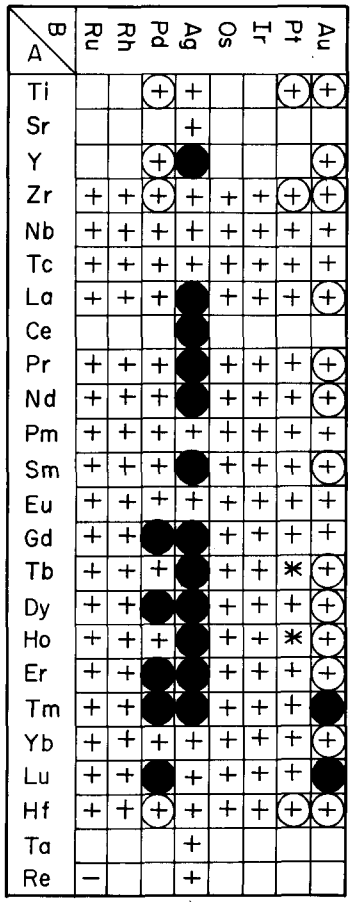

Fig. 2. Part of a two-dimensional plot illustrating the prediction of Heusler phases with the composition AB_2In: +, formation of compound with the particular type of crystal structure is predicted; −, formation of compound with the particular type of crystal structure is not predicted; ⊕, compound with the particular type of crystal structure was synthesized and appropriate information was used in the machine learning process; ⊖, compound with the particular type of crystal structure does not exist under normal conditions and this information was used in the machine learning process; ●, predicted formation of compound with the particular type of crystal structure which was confirmed by experiment; ★, predicted formation of compound with the particular type of crystal structure which was not confirmed by experiment.

The Institute of Metallurgy of Russian Academy of Sciences is now building the database of inorganic compound properties. This database contains the properties of the ternary compounds [16–18].

4.1. Database of inorganic compounds

The database containing the properties of inorganic compounds is a collection of interrelated data files intended for the use by both scientists and engineers. The database in operation contains the information about the properties of known ternary compounds (*i.e.* of compounds that comprise three chemical elements).

4.1.1. Phase subsystem

The database master file (the "phases" subsystem) contains the fundamental properties of the phase. This is a document that contains brief information about the ternary system (the temperature of the studied isothermic sections, designation of known quasi-binary sections, quantity of the ternary compounds etc.), as well as the following information about the ternary compounds: (1) qualitative and quantitative composition of the compound; (2) type of melting (congruent or incongruent) at 1 atm; (3) melting point (at 1 atm); (4) boiling point (at 1 atm); (5) decomposition temperature in the solid or gaseous state (at 1 atm); (6) type of crystal structure (the information about each crystal modification is as follows: (1) temperature and pressure to be exceeded to ensure the formation of the particular crystal modification; (2) syngony; (3) space group; (4) Z (number of formula units in unit cell)). In addition, each crystal modification of the superconducting compounds requires the information that belongs to the "superconductors" subsystem as indicated below: (1) critical temperature of transition to the superconducting state; (2) upper critical magnetic field at 4.2 K.

4.1.2. Property subsystem

The concept of an inorganic property database implies that apart from the information on the fundamental properties, additional information is stored on phases which have practical value. At present, this additional information covers ternary compounds which possess acousto-optical, electro-optical and non-linear optical properties. This particular subsystem stores the following data on compounds with definite composition and point group symmetry and possession of the above properties: (1) indices of refraction depending on the radiation wavelength; (2) coefficients of thermal expansion depending on temperature; (3) coefficients of thermal conductivity depending on temperature; (4) dielectric coefficients depending on the frequency of the electric field applied; (5) electro-optic coefficients depending on the radiation wavelength; (6) non-linear optical coefficients depending on the radiation wavelength; (7) dielectric losses (the tangent of the dielectric losses depending on the frequency of the electric field applied); (8) transmission band; (9) angle of synchronism; (10) solubility in water at room temperature; (11) heat capacity depending on temperature; (12) piezoelectric coefficients; (13) elastic constants; (14) elasto-optic coefficients depending on the radiation wavelength; (15) elastic wave velocity depending on the direction and type (longitudinal or transverse) of the elastic wave; (16) damping factors of the elastic wave in a crystal at a fixed frequency, depending on the direction and type of the elastic wave; (17) lattice parameters; (18) density; (19) Curie point.

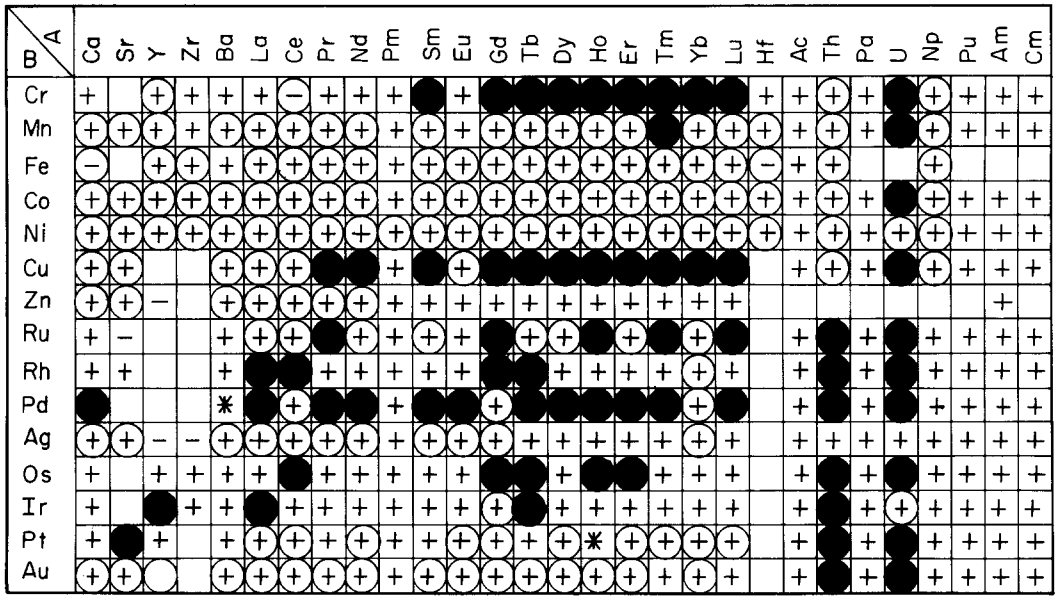

Fig. 3. Part of two-dimensional plot showing the comparison of the predicted phases of composition AB_2Si_2 with the $ThCr_2Si_2$ structure type: ○, predicted absence of compounds with the particular type of crystal structure was not confirmed by experiment; see Fig. 2 for other symbols.

4.1.3. Bibliographic subsystem

This subsystem contains the references to the published literature from which the properties and other experimental data have been extracted. This subsystem, called "References" contains the following information: (1) author's name; (2) journal reference; (3) title of the article; (4) reference to the abstract bulletin.

At present this database contains information on about 35 000 ternary compounds taken from 9000 publications. The collection and evaluation of data on the properties of binary and quaternary compounds is in progress. Access to and use of this system is menu based.

4.2. Predicting subsystem

The new perspectives of our studies in this field involve the development of a predicting subsystem based on the cybernetic predicting system as discussed above. This subsystem employs a knowledge base, a conversational processor and a monitor.

4.2.1. The knowledge base

This stores the multidimensional criteria (classification rules) already obtained for various classes of inorganic compounds for their use in the prediction of phases and forecasting. The knowledge base data are represented in the form of semantic networks or of their equivalent conjunctive–disjunctive logical expressions.

4.2.2. Conversational processor

This manages the conversation of the user with the information-predicting system, as well as giving expert function support in the given application domain.

4.2.3. The monitor

This controls the computation process and provides the interface between the functional subsystems as well as the telecommunication access to the system. In addition, the monitor signals if new experimental data contradict the predicted multidimensional criteria. Such contradictions will be eliminated by including the new data in the "learning step" and modifying the multidimensional criteria in the knowledge base.

5. Operation of the information-predicting system

The information-predicting system operates as follows (Fig. 4). The user requests the information about an existing compound of certain composition. If the data about this phase are stored in the database they can be extracted and used for further studies. If no information about the compound is stored in the database, or if the information available is incomplete, the computer in response to the user's request determines whether the regularity corresponding to the desired property for a compound of a certain type is present in the knowledge base. If the phase is present, the database supplies the appropriate set of the component properties to ensure the prediction of the desired characteristic. If the knowledge database does not have

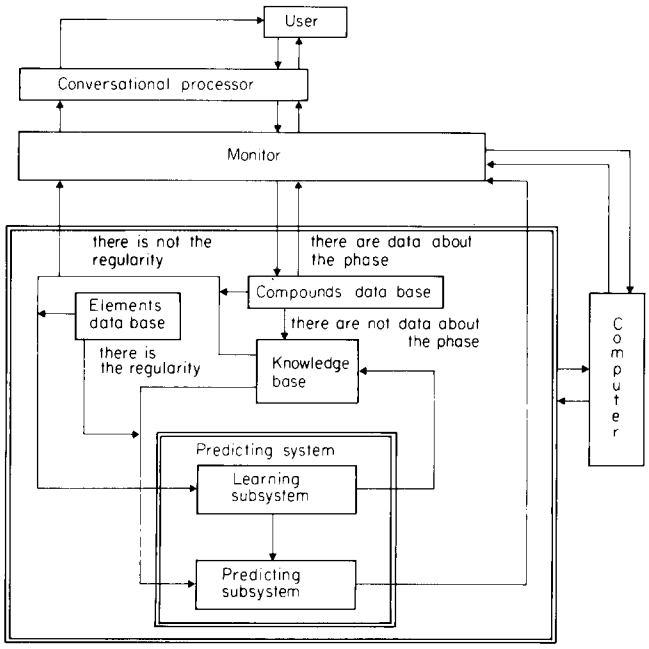

Fig. 4. Schematic diagram of an information-predicting system.

the desired regularity then the examples for the machine learning process are searched in the database. The correctness and representation of these results are estimated once more by the user and, if the sample is found adequate for machine learning, the learning and prediction subsystems process them in turn. The resultant prediction is received by the user while the classifying regularity thus shaped becomes stored by the knowledge base. The above example is the simplest of the problems than can be solved by the information-predicting system. A more complicated problem would be, for example, predicting all possible phases in ternary and multicomponent systems combined with the estimation of their properties. If the previous problem has a time complexity of $O(n)$, the latter problem has a time complexity much greater than $O(n)$.

The principles underlying the information-predicting system at present developed have been tested successfully [13] on the prototyping system supported by the BESM-6 second-generation computer.

This information-predicting system is a modified version of expert systems which find wide-ranging application both in science and in other problem-solving arenas. Such systems are widely used in chemistry, particularly in organic chemistry. The first expert systems SYNCHEM [19] and DENDRAL [20] were developed for the use in planning organic synthesis and interpreting spectral data about organic compounds. In inorganic materials science work in this field has just started. In France and Belgium there are expert systems for solving the problems of corrosive resistance of metals and alloys [21–23].

Our information-predicting system differs from conventional expert systems [19–23, 24] in that it employs an unusual procedure for receiving the knowledge from the experts and its subsystem of explanations is of a rudimentary nature. From our experience we find that any attempt to make a chemist offer an adequate explanation of the causes of the formation of one phase or another, or the nature of some properties of a compound, usually ends in failure. Instead, the chemist prefers to estimate the authenticity of data concerning the existence of the given compound or the values of the particular phase properties. While allowing for the specific features of this application domain we abandoned the idea of making the chemist outline the rules for the formation of inorganic substances with desired properties and decided to make use of the chemist's expert estimation of the data intended for machine learning.

6. Conclusions

In this process of automating certain aspects of scientific research ranging from the development of the databanks to the building of expert systems, the historical process of cognition is repeated: from collection and processing of the empirical source data to the generalization of the experimental facts. The latter is used as a basis for constructing the scientific theories reflecting the fundamental relations and correlations between the processes and phenomena studied. The development of artificial intelligence systems, such as information-predicting or expert systems, is indicative of the advent of a new type of modelling of cognitive activity [24], or knowledge engineering. This type of modelling will play an important role in the fields of science and technology where mathematical modelling and computer-aided experimentation prove to be inadequate (in chemistry and materials science).

References

1 E.M. Savitskii, N.N. Kiselyova and T. M. Shkatova, *Izv. Akad. Nauk SSSR, Neorg. Mater., 18* (1982) 235.
2 E.M. Savitskii, Yu.V. Devingtal and V.B. Gribulya, *Dokl. Akad. Nauk SSSR, 178* (1968) 79.
3 Yu.P. Vorob'ov, A.N. Men and V.B. Fetisov, *Raschyot i Prognozirovanie Svoystv Oxidov,* Nauka, Moscow, 1983.
4 V.M. Vozdvishenskij, *Prognoz Dvojnich Diagram Sostoyaniya,* Metallurgya, Moscow, 1974.
5 P. Villars, *J. Less-Common Met., 119* (1986) 175.
6 D. Pettifor, *Phys. Rev. Lett., 42* (1979) 846.
7 M.T. Yin and M.L. Cohen, *Solid State Commun., 38* (1981) 625.
8 V.P. Gladun, *Evristicheskij Poisk v Slaoshnih Sreda,* Naukova Dumka, Kiev, 1977.

9 O.Yu. MArkin, *Upr. Syst. Mash., 2* (1988) 88.
10 E.M. Savitskii, V.B. Gribulya, N.N. Kiselyova, M. Ristich, Z. Nikolich, Z. Stoyil'kovich, M. Zhivkovich and I.P. Arsent'eva, *Progonozirovante y Materialovedenii s Primenenoem EVM,* Nauka, Moscow, 1990.
11 E.M. Savitskii and N.N. Kiselyova, *Izv. Akad. Nauk SSSR, Neorg. Mater., 15* (1979) 1101.
12 E.M. Savitskii, V.B. Gribulya and N.N. Kiselyova, *J. Less-Common Met., 72* (1980) 307.
13 N.N. Kiselyova and G.S. Burkhanov, *Izv. Akad. Nauk SSSR, Met., 1* (1989) 218.
14 N.N. Kiselyova and E.M. Savitskii, *Izv. Akad. Nauk SSSR, Neorg. Mater., 19* (1983) 489.
15 E.M. Savitskii, V.B. Gribulya and N.N. Kiselyova, *CALPHAD, 3* (1979) 171.
16 E.M. Savitskii, N.N. Kiselyova, B.N. Pishik, N.V. Kravchenko and N.S. Golikova, *Dokl. Akad. Nauk SSSR, 279* (1984) 627.
17 M.S. Golikova, G.S. Burkhanov, N.N. Kiselyova, B.N. Pishik and V.B. Gribulya, *Izv. Akad. Nauk SSSR, Neorg. Mater., 25* (1989) 700.
18 N.V. Kravchenko, G.S. Burkhanov, N.N. Kiselyova, B.N. Pishik and V.B. Gribulya, *Izv. Akad. Nauk SSSR, Neorg. Mater., 27* (1991) 164.
19 H.L. Gelernter, A.F. Sanders, D.L. Larsen, K.K. Agarwal, R.H. Boivie, G.A. Spritzer and J.L. Searlman, *Science, 197* (1977) 1041.
20 D.A. Waterman, *A Guide to Expert Systems,* Addison-Wesley, New York, 1986.
21 S. Audisio and M.A. Audisio, *Mater. Technol., 6* (1989) 49.
22 J. Schreurs and N. Pessall, *ICTTE 86: Technology Shaping the Future, Proc. Int. Congr. on Technology and Technology Exchange, 1986,* p.9.
23 W. Bogaerts and M. Vancoille, *Bull. Cercle Etud. Met., 15* (1988) 5/1.
24 G.S. Pospelov, *Iskustvennij Intellekt — Osnova Novoy Informatzionnoj Technologii,* Nauka, Moskow, 1988.