ELSEVIER

Journal of Alloys and Compounds 279 (1998) 8–13

Journal of ALLOYS AND COMPOUNDS

Computational materials design using artificial intelligence methods

N.N. Kiselyova[a,*], V.P. Gladun[b], N.D. Vashchenko[b]

[a]*A.A. Baikov Institute of Metallurgy, Russian Academy of Sciences, Leninskii Prospect, 49, 117334 Moscow, Russia*
[b]*Institute of Cybernetics, National Academy of Sciences of Ukraine, Prospect Acad. Glushkova, 40, 252650, GSP, Kiev-22, Ukraine*

Abstract

Problems associated with the design of new materials can be divided into the search for new materials with predefined properties, and the search for optimum processing conditions for the production of new materials. We use artificial intelligence methods for solution of the first problem, which requires knowledge of values of the elements' properties for prediction. The comparison of predictions with experimental data shows that the average reliability of the prediction exceeds 80%. Several tables of predicted compounds are presented relative to predictions of the crystal structure type $ThCr_2Si_2$ for compounds with composition AB_2Si_2 and $A^{II}B^{III}F_5$. For the solution to the second problem, we briefly discuss two methods we have used for optimizing the process for producing materials which have been predicted.

Keywords: Pyramidal nets; Materials design; Materials property prediction; Artificial intelligence

1. Introduction

Before speaking about the problem of computer design of materials, we shall dwell on conventional methods applied to the search for new inorganic substances. The first method is casual discovery of new properties applied frequently to well known substances. For example, in 1911 Kamerlingh Onnes obtained the effect of superconductivity in mercury, Kerr obtained a quadratic electro-optical effect in 1875, and later Rontgen and Kundt a linear electro-optical effect in quartz and tourmaline. This approach is similar to the 'trial-and-error' method. The second method has a more rigorous scientific basis and is connected with the search for analogs of known compounds for the purpose of improvement of their parameters. The scientific basis of such searches is a paradigm which dominates modern science, e.g., conservation laws, concepts about an intrinsic structure of substances and bond types of atoms, as well as the set of empirical and semi-empirical knowledge accumulated by chemistry and materials science. Unfortunately, modern physics gives in a form suited for chemists only the paradigm, but not the methods of new substance calculation. Quantum mechanical calculations, although striking to the chemist's imagination, are very difficult in the case of practical new substances. As a result, many chemists continue to use only their knowledge and experience for developing substances.

What is this knowledge and experience of a chemist? First, it contains a set of experimental facts (property values of substances, process parameters, etc.). Second, it contains regularities connecting and/or systematizing these facts and determining the ways and limitations for application of either method used in chemical practice. Mendeleev's periodic law is a most remarkable example of knowledge of the second kind. Translated into computer language, it is possible to consider a database as a repository of knowledge of the first kind (facts) and a knowledge base, a compulsory component of an artificial intelligence or 'expert' system, which is quite appropriate for the storage of regularities and rules. This expert system approach, referred to as computer design, is focused on new inorganic materials and simulates a conventional way of searching for new materials with predefined properties where the new materials are analogs to materials already known. In this paper, the development of an expert system for the design of new inorganic materials, its components, and preliminary results achieved by realization of the proposed ideas are presented and discussed.

2. Databases

A materials database (DB) is a critically important component of any system for inorganic materials design. It can be a part of a larger DB or an independent information system. It is important in calculations that high quality data

*Corresponding author.

0925-8388/98/$19.00
PII: S0925-8388(98)00606-9

be used, which requires expert assessment of the stored information. Although this introduces an additional problem in finding such experts, the reliability of experimental data is increased significantly. It should be noted that the expert assessment of data is a feature of all DBs developed by us. The DBs were developed for IBM PC computer. The most interesting of them are the following:

1. An inorganic ternary compound properties DB was built by us in the 1970's [1,2]. It contains information on more than 37 000 ternary compounds taken from more than 10 000 publications. Some of the data has been assessed by materials experts.
2. A phase diagram DB of material systems with intermediate semiconducting phases contains information on physical–chemical properties of the intermediate phases [3] and the most important Pressure–Temperature–Concentration phase diagrams of semiconducting systems evaluated by qualified experts. Currently the DB contains information on several tens of semiconducting systems.
3. An acousto-, electro-, and nonlinear-optical properties [4–6] DB which contains information on crystals evaluated by experts. In addition, DB includes extensive graphical information about the properties of the materials. This version of the DB was developed for platform independence using DBMS Oracle. In addition to a DB in Russian, a version in English also has been developed.

3. Search for regularities and prediction of new materials

The search subsystem for discovering regularities in the information stored in a DB is the most complicated part of a system for computer design of materials. The problem of new materials design can be divided into two parts: (1) the search for new substances with predefined properties, and (2) development of optimum conditions for the fabrication of new materials. The problem of the search for new materials also has two aspects: (1a) predicting composition of the new materials and their properties based on knowledge of properties (chemical elements or simple compounds), and (1b) the choice of substances that best correspond to certain requirements. All our attempts to extract rules connecting the formation of a certain compound to properties of its constituent elements from chemists or material scientists were unsuccessful.

Chemists could not even indicate the list of those parameters of chemical elements which determined, e.g., the possibility of forming binary compounds of given composition. In the best case they indicate the list of the traditional parameters of elements that is limited by atomic (ionic) radii, electronegativities, electronic concentration (ratio of number of valence electrons to number of atoms in the compound) or functions of their parameters (well known Darken–Gurry rules for the prediction of metal systems with continuous solid solutions [7], Mathias rule for the prediction of superconductors with A-15 crystal structure [8], Hume-Rothery rules for the prediction of the solubility of metals in the solid state [9], Goldschmidt rule for the prediction of new oxides with perovskite structure [10], QSD-method for the prediction of new ferroelectrics with ABO_3 composition and other substances [11], and so on). These elemental properties are important but this list is not limited by these parameters of the elements for prediction in inorganic chemistry, i.e. the choice of these properties is only a tradition.

In addition, it is impossible to speak about the importance of an individual property in the whole range of its values. As a rule some property is of importance for classification in only a certain range and for the substances of a certain kind. In most cases, chemical phenomena are determined by a set of properties and thus, it makes sense to speak only about importance of set intervals of properties. The search for such multi-dimensional regularities is a complicated problem for chemists. Therefore we concluded that the process of search for multi-dimensional regularities connecting properties of substances with component parameters should be executed by a computer. A typical problem for chemists is the assessment of initial data and the subsequent interpretation of observed regularities. In addition, it would be desirable if the chemist had generated an initial set of component properties. In this case the computer should choose the most important features from the properties.

4. CONFOR – a classification and prediction system

We use the methods of artificial intelligence (AI), more specifically, machine learning based on examples of already known substances in the search for regularities. The method is based on concept formation ideas [12–15] and involves the following processes:

- Representation of initial information about known physical–chemical systems having predefined properties with the use of pyramidal networks;
- discovery of regularities which define classes of materials with known physical–chemical systems;
- representation of these regularities as a set of logical expressions;
- prediction of new materials having predefined properties based on knowledge of constituent element properties.

The herein referred to physical–chemical system is a material system, e.g., compound, solid solution, or heterogeneous mixture, which is formed from chemical elements. A class of physical–chemical systems is a set of

materials having a common target property such as compound formation, a particular composition, or compounds with a definite crystal structure type.

A method for discovering such target properties is pyramidal networks (special acyclic oriented graphs) which allow the search for new materials to be accomplished. The method of regularities discovery and prediction on the basis of pyramidal networks is exemplified in a software application called CONFOR (CONcept FORmation). Prediction is performed by determining class membership which first requires identifying materials knowledge that characterizes the class of physical–chemical systems. It is often necessary to characterize a class through an analysis of known physical–chemical systems which possess a desired property (denoted as the 'in-class') together with systems which do not (denoted as the 'out-of-class') [12–15]. Once CONFOR is able to distinguish between these classes, materials with similar physical–chemical systems, for which no values of these properties are known, are then evaluated. This process in essence is a learning or 'training-and-testing' process. The set of physical–chemical systems used to characterize the class of interest is referred to as a learning set.

The knowledge used for prediction is the conjunction of attribute values that characterize the desired materials to include the desired property attribute values, as well as other attributes, and particular values for those attributes that do not exist together with the desired property values. This knowledge can be described by a logical expression in which essential combinations of attribute values are represented by a conjunction of variables which designate the in-class. The logical expression describing an in-class of physical–chemical systems in essence is a logical model of the class. Formation of logical models for classes of physical–chemical systems via search-driven analysis of learning sets is often referred to as 'knowledge discovery', 'knowledge mining' and 'concept learning'. The last term defines the process most adequately. In this case the concept is a generalized model of some class of physical–chemical systems.

After building a logical model for some class of physical–chemical systems, prediction of physical–chemical system members of this class is reduced to comparison of their attribute descriptions with the logical expression defining the class. The comparison is performed by calculation of the logical expression value after substitution of '1' for variables that are available in the description of material and '0' for other variables. If the value of the logical expression equals '1' the tested physical–chemical system is determined to form or possess a certain property of interest.

5. Use of CONFOR for new compound prediction

The physical–chemical basis of our approach is the Mendeleev law. According to this law the periodic change in the properties of chemical systems depends on the properties and nature of the elements which form these systems (compounds, solutions and so on). Mendeleev's law is an expression of the condition of compactness for compound classes in the multi-dimensional property space of the chemical elements.

The use of CONFOR for predicting new inorganic

Table 1
Characteristics of the prediction results for the inorganic phases

Compounds/Systems	Characteristics to be predicted	Experimental tests for March 1997	Error of prediction, %
ABX (X=Se,Te)	Compound formation	92	42
ABX_2 (X=O,S,Se,Te)	Compound formation	316	10
ABX_3 (X=O,F,S,Cl,Se,Br,Te,I)	Compound formation	368	13
ABX_4 (X=O,F,Cl,Br,I)	Compound formation	389	5
A_2BX_2 (X=S,Se)	Compound formation	22	9
AB_2X_4 (X=O,F,S,Cl,Se,Br,Te,I)	Compound formation	727	16
$A_2B_2X_7$ (X=O,S,Se)	Compound formation	97	26
$A(Hal)_2$–B(Hal)	Systems w/ compounds	101	8
AB_2X_4 (X=O,S,Se,Te)	Structure type	359	6
ABX (X=Al,Si,P,Ga, Ge,As,Pd,In,Bi)	Structure type	42	52
ABO_3	Perovskite structure	180	11
$A_2B_2O_7$	Pyrochlore structure	71	18
AB_2X_2 (X=Al,Si,P,Ge,As,Sb)	$ThCr_2Si_2$ structure	151	7
ABX_2 (X=Al,Co,Ni, Cu,Ga,Pd,In)	$MnCu_2Al$ structure	52	13
$A_x(SO_4)$*$B_z(SO_4)_w$	Compound formation 1:1	127	4
$A(NO_3)_x$*$B(NO_3)_y$	Compound formation 1:1		
$ABDO_4$	Compound formation	22	3
			Average=15%

Table 2
Part of predictions of the crystal structure type $ThCr_2Si_2$ for compounds with composition AB_2Si_2

A \ B	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ru	Rh	Pd	Ag	Os	Ir	Pt	Au
Ca	+	⊕	↔	⊕	⊕	⊕	⊕	+	+	©	⊕	+	+	+	⊕
Sr		⊕		⊕	⊕	⊕	⊕	-	-		⊕			©	⊕
Y	⊕	⊕	⊕	⊕	⊕		-				-	+	©	+	Ø
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	+	+		+	+	+		+	+	+	+	+	+	+	+

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 crystal structure $ThCr_2Si_2$ was synthesized and appropriate information was used in the computer learning process;
↔ compound with the crystal structure $ThCr_2Si_2$ does not exist under normal conditions and this information was used in the computer learning process;
© predicted formation of compound with the crystal structure type $ThCr_2Si_2$ which was confirmed by experiment;
⊗ predicted formation of compound with the crystal structure type $ThCr_2Si_2$ which was not confirmed by experiment;
Ø predicted absence of compound with the crystal structure $ThCr_2Si_2$ which was not confirmed by experiment; empty square – indeterminate result.

compound crystal structures shows [16–20] that the average reliability of the prediction, based only on constituent property knowledge of chemical elements or simple compounds, exceeds 80% (Table 1). The shaded cells of Table 2 show the results of comparing the predicted new phases of composition AB_2Si_2 possessing crystal structure type $ThCr_2Si_2$ with subsequent experimental data. These compounds are of some interest in the search for new magnetic materials.

Of the 77 predictions, 73 compounds agreed with experimental data. Listed in Table 3 are the results of predicting the formation and crystal structure type, at room temperature and atmospheric pressure, of compounds with the composition ABF_5 [19]. Because of this good agreement, these compounds are useful in the search for new electro-optical materials.

What properties of the constituent components do we use for the description of physical–chemical systems? First, all fundamental properties of chemical elements: the distribution of electrons in the energy levels of isolated atoms, ionization potentials, ionic, covalent or atomic radii of elements (cations), melting points, standard entropy of individual substances and the like. Second, all properties of simple compounds – oxides, chalcogenides, halides, etc. – as required by the composition of the compounds predicted.

Now we develop the information-predicting system [16,17] to be used for automating the prediction of the 'new compound'. The system includes a database, a knowledge base, and the predicting system based on the program CONFOR [12–15]. Our approach has made it possible to solve problems of the following types [16–20]:

- prediction of compound formation or non-formation for ternary systems;
- prediction of the possibility of forming ternary and more complicated compounds of desired composition;
- prediction of phases with defined crystal structures;
- estimation of phase quantitative properties (critical temperature of transition to superconducting state, homogeneity region, etc.).

6. Conclusions

We have applied a search-based method, AI, for discovering regularities in the design of new inorganic materials. The effectiveness of the proposed approach is illustrated in Table 1, 2 and 3. Aside from the prediction accuracies across various crystal structure types in Table 1, new magnetic and electro-optical materials with specific crystal structures have been predicted.

To further improve the processing time and accuracy of CONFOR for predicting new inorganic materials, there are aspects of the method which will require further development. One very significant aspect is that of discretization

Table 3
Part of predictions of the crystal structure type for compounds with composition $A^{II}B^{III}F_5$

A/B	Be	Ca	V	Fe	Ni	Cu	Zn	Sr	Cd	Sn	Ba	Sm	Eu	Yb	Hg	Pb
B							?	*				B				
Al		F		G		C	G	(B)	?	?	(G)	B	(B)		G	(B)
Sc		(*)				G		*	(*)		*				*	(*)
Ti	?	©	?	C	C	-	?	(S)			(B)	C			B	
V	B	©				B	B	(S)			(B)				B	?
Cr	C	©			?	C	?	(B)	(F)	F	(G)				?	
Mn				F			F		F		(G)	B			G	
Fe	F	(F)			F	F	F	(S)		F	(B)		*	*	B	
Ga	C	©	C	C		?	C	(B)	(F)		(G)				C	(B)
As			S					*					*	*	*	
Lu		(*)						(*)				B	B	B		*
Tl	*		S					*	*		(B)				*	
Ac		C	S				?	*	*		*		B	B		

Designations:
B – formation of compound with the crystal structure type $BaFeF_5$ is predicted;
C – formation of compound with the crystal structure type $CaCrF_5$ is predicted;
G – formation of compound with the crystal structure type $BaGaF_5$ is predicted;
F – formation of compound with the crystal structure type $CaFeF_5$ is predicted;
S – formation of compound with the crystal structure type $SrFeF_5$ is predicted;
- the crystal structure differing from those listed above is predicted;
* prediction of no compound with composition ABF_5;
(B), ©, (G), (F), (S) – compound with corresponding type of crystal structure was synthesized and appropriate information was used in the computer learning 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 learning process;
(*) compound ABF_5 is not formed and this fact was used in the computer learning process;
? or empty square – indeterminate result.

or binning. Although this was briefly presented, the optimum selection of ranges or bins for attribute values is an open research topic. In addition, when prediction accuracy is poor, as illustrated in Table 1 for material systems such as ABX (X=Al, Si, P, Ga, Ge, As, Pd, In, Bi), a need exists to help identify the cause of the poor prediction accuracy. Such causes range from missing information, i.e., attributes regarding structure and/or formation which are required to distinguish between in-class and out-of-class membership, to nonlinear relations between attributes, which if known and properly represented would again improve in-class and out-of-class discrimination.

The computer design of inorganic materials is not limited to the search for new compounds with predefined properties. Surely, information beyond the structure and properties of compounds can be organized into databases to support the next step, which is the search for the optimum process technology and conditions. At this stage, without the availability of such process data we propose to complement the use of AI with the use of various mathematical methods for optimization, enabling sharply reduced time and costs in the development of new materials.

The optimization of processes for the production or treatment of materials, which were predicted by the first stage, directs, first of all, the search for extrema of the target properties of interest. It is necessary to use the wide spectrum of mathematical optimization methods for the solution to this problem. The solution can be a theoretical one or it can include an experiment that is carried out according to a formal plan. The information-calculating system, developed at the Baikov Institute, is an example of the first approach. Using information from a DB [6] and an algorithm based on the Lagrange factors methods, it enables the calculation, in anisotropic acousto-optical, electro-optical, and nonlinear-optical crystals, of the crystallographic direction which corresponds to extreme values of the target property. We carried out the second approach using statistical methods of design of experiments for the search for new superconducting materials [20] also.

Acknowledgements

The authors thank Drs. S.R. LeClair and A.G. Jackson for collaboration in this work.

References

[1] E.M. Savitskii, N.N. Kiselyova, B.N. Pishik, et al., Dokl. Akad. Nauk SSSR 279 (1984) 627.
[2] N.N. Kiselyova, N.V. Kravchenko, V.V. Petukhov, Inorg. Mater. 32 (1996) 567.
[3] V.S. Zemskov, N.N. Kiselyova, et al., Inorg. Mater. 31 (1995) 1096.
[4] M.S. Golikova, G.S. Burkhanov, N.N. Kiselyova, et al., Izv. Akad. Nauk SSSR, Neorg. Mater. 25 (1989) 700.
[5] N.V. Kravchenko, G.S. Burkhanov, N.N. Kiselyova, et al., Izv. Akad. Nauk SSSR, Neorg. Mater. 27 (1991) 164.
[6] N.V. Yudina, V.V. Petukhov, B.A. Cheremushkin, et al., Crystallogr. Rep. 41 (1996) 464.
[7] V.M. Vozdvizhenskii, Prediction of binary phase diagrams, Metallurgia, Moscow, 1975, p. 223
[8] B.T. Mathias, Phys. Rev. 97 (1955) 74.
[9] W. Hume-Rothery, G.V. Raynor, Proc. R. Soc. A174 (1940) 471.
[10] V.M. Goldschmidt, Skifer utgitt av del Norske Videnskaps. Akademii Oslo, I Matem. Natur-vid Kiass 2 (1926) 97.
[11] K.M. Rabe, J.C. Phillips, P. Villars, I.D. Brown, Phys. Rev. B 45 (1992) 7650.
[12] V.P. Gladun, Processes of Formation of New Knowledge, SD 'Pedagog 6', Sofia, 1995, p. 192.
[13] V.P. Gladun, Z.L. Rabinovich, in: Machine Intelligence, 9, Ellis Herwood Ltd., Chichester, 1980, p. 299–309.
[14] V.P. Gladun, Computers and Artificial Intelligence, Bratislava, No's. 1–5 (1986) 13.
[15] V.P. Gladun, N.D. Vashchenko, Cybern. Sys. Anal. 31 (1995) 207.
[16] E.M. Savitskii, V.B. Gribulya, N.N. Kiselyova et al., Prediction in Material Science Using Computer, Nauka, Moscow, 1990, p. 86.
[17] N.N. Kiselyova, J. Alloys and Comp. 197 (1993) 159.
[18] E.M. Savitskii, N.N. Kiselyova, Izv. Akad. Nauk SSSR, Metally. 1 (1984) 191.
[19] N.N. Kiselyova, Zh. Neorg. Khim. 42 (1997) 896.
[20] E.M. Savitskii, N.N. Kiselyova, T.M. Shkatova, Izv. Akad. SSSR, Neorg. Mater. 18 (1982) 235.