PERGAMON

Engineering Applications of Artificial Intelligence 13 (2000) 533–542

Engineering Applications of
ARTIFICIAL
INTELLIGENCE

www.elsevier.com/locate/engappai

Databases and semantic networks for the inorganic materials computer design

N.N. Kiselyova

A. A. Baikov Institute of Metallurgy and Materials Science, Russian Academy of Sciences, Leninskii Prospect, 49, 117334 Moscow, Russia

Abstract

At present, hundreds of databases (DBs) on substance and material properties are being developed. The prime aim of their operation is information service. Fifteen years ago, we proposed to use an extensive information of DBs not only for information service, but also for searching regularities in data and the application of these regularities for the prediction of new substances. The semantic networks of special kind were used for the search for regularities. Using deduced regularities we have predicted thousands of new compounds in ternary, quaternary and more complicated systems, and estimated some of their properties (crystal structure type, melting point, homogeneity region etc.). The comparison of our predictions with experimental data, obtained later, showed that the average reliability of predicted inorganic compounds exceeds 80%. © 2000 Published by Elsevier Science Ltd.

Keywords: Database; Artificial intelligence; Computer design; Inorganic material; Inorganic compound; Crystal structure type; Semantic network; Growing pyramidal network; Prediction; Information-predicting system

1. Introduction

The idea of development of artificial intelligence (AI) in the model of '*Homo sapiens*' occurred with the advent of "clever machines" — the computers. At present, ideas of AI are transformed to the development of robots performing some of the human duties to be performed under unfavourable conditions, programs which would extend the capabilities of dialoguing with the computers, expert systems allowing the solution of tasks which defy mathematical formulation, etc. The most interesting AI applications are the data processing program systems for large symbolic information bulks. The target of this processing is to search for regularities in data. At the beginning, these programs were made for robot learning in purposeful behaviour in actual practice and as a tool for the analysis of audio- and video-information to be input into computer. But soon it became evident that the area of these programs application can be much more encompassing and that these are tools extending the human capabilities in cognition of the universe. The AI methods named as computer learning came into use widely for analyses of geological and geophysical information with the aim of prediction of deposits or earthquakes, for technical diagnosis of machinery faultinesses, for medical diagnostics, for analysis of spectral data for the purpose of detection of various chemical and physical effects, etc. This paper presents the results of computer learning applications in data processing of a great body of information about inorganic substances with the goal of prediction of new materials with the predefined properties.

2. Semantic networks for the representation knowledge about inorganic substances

Data processing is an analysis of properties and relations of objects with an aim to detect various connections in between. Therefore, the most optimal form of such a data representation in the computer memory

0952-1976/00/$ - see front matter © 2000 Published by Elsevier Science Ltd.
PII: S0952-1976(00)00033-6

is associative structures which allow to trace the connections without a great body of sorting information. The mathematical model of associative structures is a special kind of graph — the semantic networks (SNs), Fig. 1, describing the connections between objects, properties of objects and their status (Gladun, 1994). The SNs in inorganic chemistry can represent the connections between properties of components of known chemical systems and properties of these systems. It is important that the chemical systems can be divided into distinguishable classes with their physical and chemical properties. These classes were determined by some concepts, for example, "systems with formation of compounds of definite composition", "compounds with definite crystal structure type", "superconductors with the nitrogen critical temperature of transition to the superconducting state", and so on. The search for sets of component properties' intervals, which cause the system membership to be in a certain class, is one of the chemical data processing task using the SNs. The result is a general classifying regularity (a computer form of some known concept description). The process of concept formation is referred to as computer learning. The periodicity of chemical elements' properties and, following on from this fact, the periodicity of compounds' properties, allow the use of the formed computer concepts for recognition of membership of unknown chemical element set to one or other class, which are described by formed concepts. This process is called the results prediction.

3. Databases as a foundation of semantic networks building

The application of SNs, and other empirical and semi-empirical methods for finding regularities is rather put into use in case of complete and qualitative

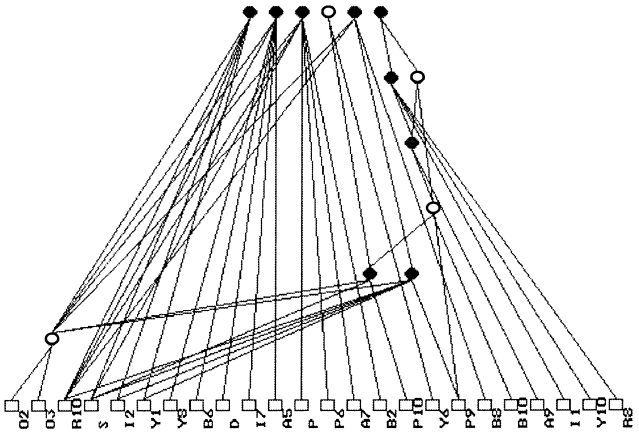


Fig. 1. Example of SN.

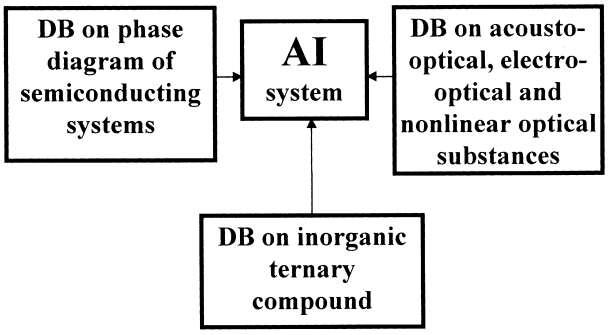


Fig. 2. Databases for the search for new inorganic compounds.

data. Our experience of SN applications, shows that the number of erroneous predictions varies proportionally with number of errors in experimental data to be processed and the reliability grows with an increase of initial data volume (reliability mounts to a limit with an increase of size and representativeness of learning set). Consequently, the application of the methods on SNs, for searching regularities implies the use of databases (DBs), containing extensive bulks of qualitative information, as a basis. With this aim in mind we develop the DBs containing data with the qualified expert assessment. The most interesting of them are: the DBs on materials for electronics with completely assessed information (Yudina et al., 1995; Zemskov et al., 1994) and inorganic ternary compound properties DB containing partially assessed information (Kiselyova et al., 1996; Savitskii et al., 1984) (Fig. 2).

1. a *phase diagram* DB of material systems with intermediate semiconducting phases (Zemskov et al., 1995) contains information on physical–chemical

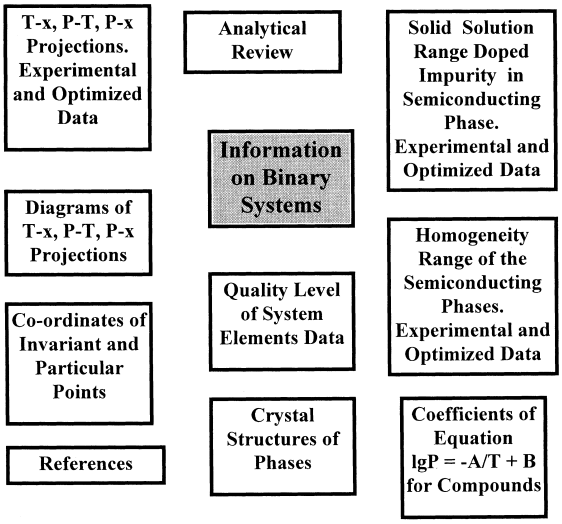


Fig. 3. Structure of DB on binary semiconducting systems.

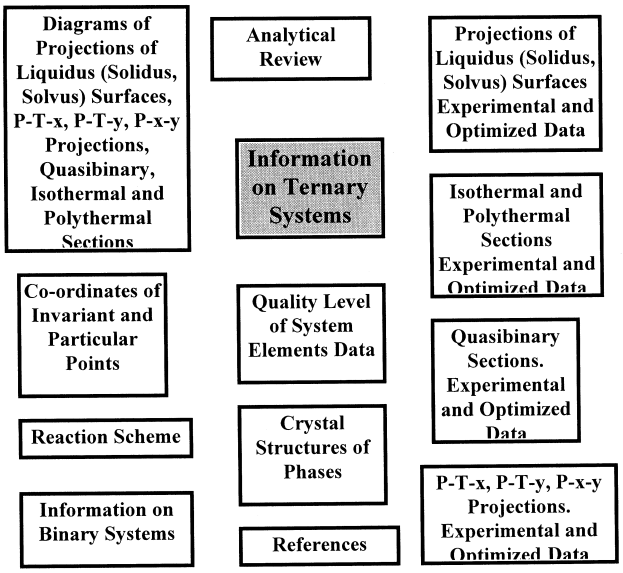


Fig. 4. Structure of DB on ternary semiconducting systems.

properties of the intermediate phases and the most important Pressure–Temperature–Concentration phase diagrams of semiconducting systems, evaluated by qualified experts (Figs. 3 and 4). Currently, the DB contains information regarding several tens of semiconducting systems.

2. DB of *acousto-, electro-, and nonlinear-optical* properties (Yudina et al., 1996) contains information regarding crystals evaluated by experts (Fig. 5). In addition, DB includes extensive graphical information about the properties of the materials.
3. A DB of *inorganic ternary compound properties* was built by us in the 1970s (Kiselyova et al., 1996; Savitskii et al., 1984). It contains information about thermochemical, crystal chemical and superconducting properties of more than 37,000 ternary compounds taken from more than 11,000 publications (Fig. 6). Some of the data have been assessed by materials experts.

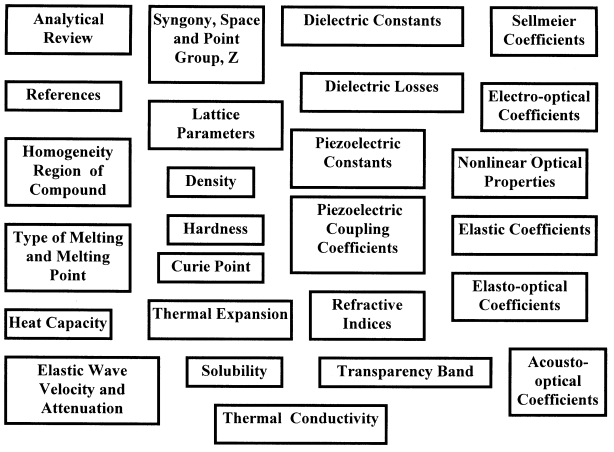


Fig. 5. Structure of database on crystals with acousto-, electro- and nonlinear-optical properties "Crystal".

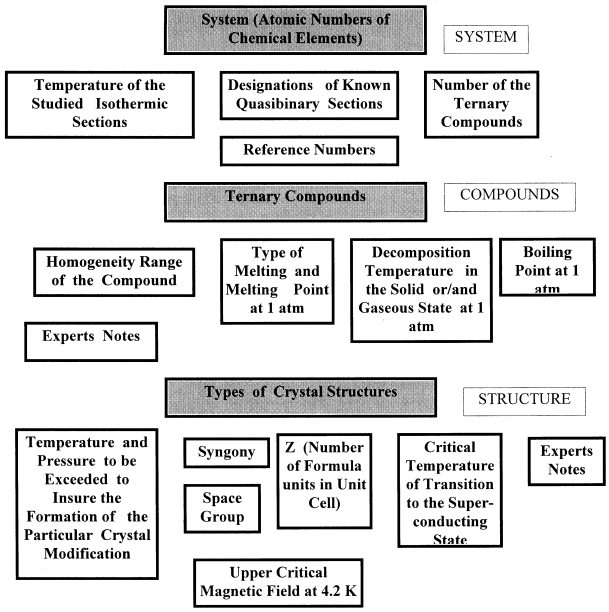


Fig. 6. Structure of database of ternary inorganic compounds "Phases".

Our investigations, aimed at predicting new materials for electronics and other applications, are based

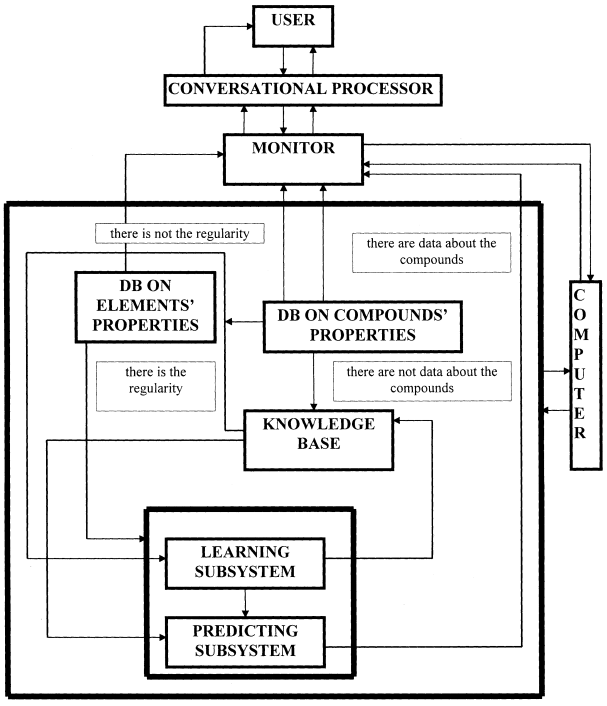


Fig. 7. Schematic diagram of an information-predicting system.

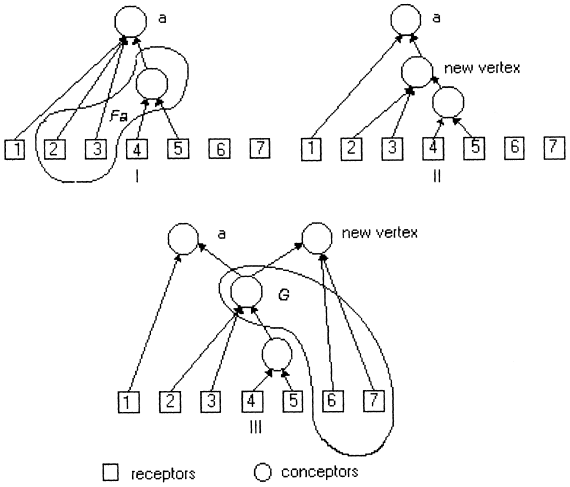
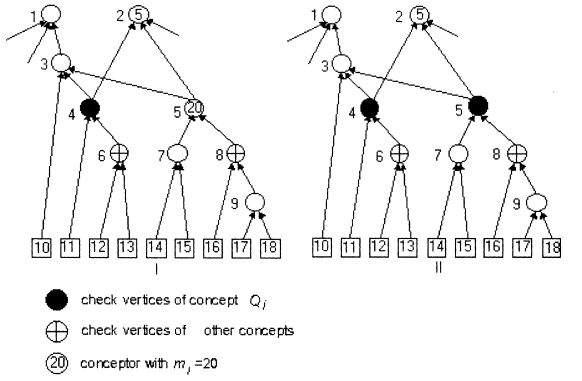


Fig. 8. GPN building process (Gladun, 1994).



Concept formation=>

$$=> \ (12 \wedge 13) \wedge \neg 11 \vee (16 \wedge 17 \wedge 18) \wedge \neg (14 \wedge 15)$$

Fig. 9. Concept formation process (Gladun, 1994).

on these DBs. The development of a DB is connected with building of some predefined SN that represents the objective interrelations between the properties of substances. DB information, however, does not provide direct answers to the connection between the substance properties and constituent component properties. The AI application makes it possible to search for such connections.

4. Application of artificial intelligence and databases to the new inorganic materials computer design

It is impossible to use the DBs completely without special software for searching for regularities in the data. During the last quarter of the century, we and our colleagues at the Glushkov Institute of Cybernetics, National Academy of Sciences of Ukraine, have been working to solve the problems of AI application in the prediction of new inorganic materials with predefined properties (Kiselyova, 1993, 1997, 1998; Kiselyova et al., 1977; Savitskii et al., 1990). These investigations aim to develop an information-predicting system (Fig. 7) for inorganic materials computer design. This system is based on SNs and DBs (Kiselyova, 1993, 1997; Savitskii et al., 1990).

We use SN with such a system of computer learning that represents information about known chemical sys-

Table 1
Comparison of predictions with new experimental data

Compounds/systems	Characteristics to be predicted	Experimental tests for January 2000	Error of prediction (%)
ABX (X = Se, Te)	Compound formation	100	44
ABX_2(X = O, S, Se, Te)	Compound formation	337	10
ABX_3(X = O, F, S, Cl, Se, Br, Te, I)	Compound formation	420	11
ABX_4(X = O, F, Cl, Br, I)	Compound formation	393	5
A_2BX_2(X = S, Se)	Compound formation	24	9
AB_2X_4(X = O, F, S, Cl, Se, Br, Te, I)	Compound formation	761	16
$A_2B_2X_7$(X = O, S, Se)	Compound formation	97	26
$A(Hal)_2$ - B(Hal)	Systems w/ compounds	108	10
AB_2X_4 (X = O, S, Se, Te)	Structure type	381	7
ABX(X = Al, Si, P, Ga, Ge, As, Pd, In, Sb, Bi)	Structure type	78	35
ABO_3	Perovskite structure	186	13
$A_2B_2O_7$	Pyrochlore structure	74	15
AB_2X_2(X = Al, Si, P, Ge, As, Sb)	Structure type	200	8
ABX_2(X = Co, Ni, Cu, Pd)	$MnCu_2Al$ structure	28	14
AB_2X(X = Al, Ga, In)	$MnCu_2Al$ structure	24	13
$A_x(SO_4)_y*B_z(SO_4)_w$ and $A(NO_3)_x*B(NO_3)_y$	Compound formation 1:1	130	4
$ABDO_4$	Compound formation	28	4
			Average = 14%

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

Fig. 10. Part of a table illustrating the prediction of compounds with the composition ABO_3. (*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 learning process; ↔ compound with composition ABO_3, does not exist under normal conditions and this information was used in the computer learning 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; empty square — indeterminate result).

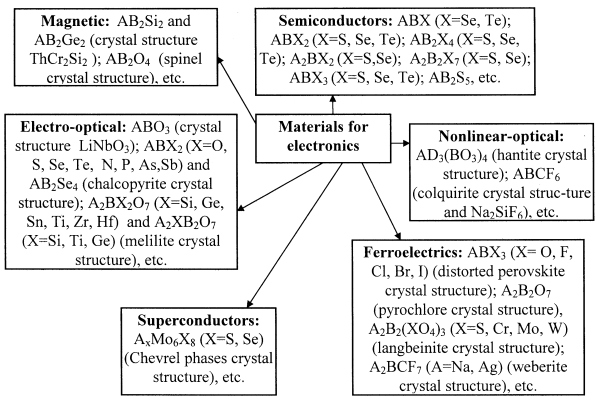


Fig. 11. Examples of classes of compounds, which prediction is realized.

tems like — growing pyramidal networks (GPNs). A pyramidal network is an acyclic oriented graph having no vertices but one entering arc. If the processes of concept formation are determined in the network then the pyramidal network is designated as a growing one (Gladun, 1994). GPN is built during the process of objects input (Fig. 8). Each object (chemical 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 nearby values of components' properties are united into one interval using a special program or the experience of a researcher. Concept formation process (Fig. 9) consists of the analysis of vertices in built network and the choice of those ones that are the most typical for each class (Gladun, 1994). These vertices became the checking vertices. The resultant concepts (classifying regularities) can be stored in

A / B	Li	Be	Al	K	Sc	V	Cr	Fe	Co	Ni	Ga	Ge	Y	Nb	Mo	Ru	Rh	Pd	Ag	In
Zn			©									-	-	-	-	-	-	-	-	
Ga	+	+		+	⊕	+	+	+	+	+		-							-	+
In	+	+		+	⊕	+	+	+	+	+	+	+	⊕	+	+	+	+	+	+	
Sn	O	-	O	-	-	-		⊕	⊕	⊕	-	-	-	↔	-	-	-	-	↔	©
Lu	-	-		-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	+
Ta		-			-	-	-	-	-	↔		-	-	↔	-	-	-	-		+
Au	-	-		-	-	-	-	-	-	-		-	-	-	-	-	-	↔	↔	+
Tl			-									-							-	+
Pb	-	-	O	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+

Fig. 12. Part of table illustrating the prediction of Heusler phases with the composition AB_2Cu. (*Designations:* + formation of compound with the crystal structure type resembling the Heusler alloys is predicted; − formation of compound with the crystal structure type resembling the Heusler alloys is not predicted; ⊕ compound with the crystal structure type resembling the Heusler alloys was synthesized and appropriate information was used in the computer learning process; © predicted formation of compound with the crystal structure type resembling the Heusler alloys, which is confirmed by experiment; ⊗ predicted formation of compound with the crystal structure type resembling the Heusler alloys, which is not confirmed by experiment; ○ predicted absence of compound with the crystal structure type resembling the Heusler alloys, which is confirmed by experiment; ↔ compound with the crystal structure type resembling the Heusler alloys does not exist under normal conditions and this fact was used in the computer learning process; empty square — indeterminate result).

computer memory and printed or read out in the form of a learned SN or an equivalent Boolean expression. These values of the component properties form the variables. During the prediction process the computer receives only 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 DB. They are substituted into the GPN and the researchers can easily obtain the necessary prediction.

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 (at 1 atm), 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 are predicted (Kiselyova, 1993, 1997, 1998; Kiselyova et al., 1977; Savitskii et al., 1990).

Our approach has made it possible to solve problems of the following types (Kiselyova, 1993, 1997, 1998; Kiselyova et al., 1977; Savitskii et al., 1990) (Table 1):

- 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 systems).

Predicted compounds were then searched for new magnets, semiconductors, superconductors, electro-optical, acousto-optical, nonlinear-optical and other materials required for new technologies (Kiselyova, 1993, 1997, 1998; Kiselyova et al., 1977; Savitskii et al., 1990). The comparison of these predictions with the experimental data, obtained later, showed that average reliability of predicted compounds exceeds 80% (Table 1).

In Fig. 10 the comparison between the results after predicting the compounds with composition ABO_3 (Kiselyova et al., 1977) and the new experimental data is illustrated. It was the first prediction which we carried out 25 years ago. Only one prediction was detected to be erroneous ($CuCeO_3$). These compounds were predicted in the process of the search for new ferroelectrics. More recently, we predicted formation of other compounds of this kind and other substances for electronics (Fig. 11).

A^{II} B^{III}	Be	Mg	Ca	Ti	V	Mn	Fe	Co	Ni	Cu	Zn	Ge	Sr	Cd	Ba	Hg	Pb
B	↔		⊕		+	⊕		↔	↔	⊕			⊕	©	⊕		↔
Al	⊕	⊕	⊕		©	⊕	⊕	⊕	⊕	⊕	⊕	↔	⊕	⊕	⊕	↔	⊕
Sc		⊕	⊕	+	+	+	+	+	+	↔	©	+	©	©	⊕	+	
Ti		©	©	+	+	⊕	©	+	+	+	+		⊕	+	©	+	+
V		⊕	⊕	+	+	⊕	⊕	⊕	+	+	⊕		⊕	⊕	©	©	+
Cr	⊕	⊕	⊕	+	+	⊕	⊕	⊕	⊕	⊕	⊕	+	⊕	⊕	⊕	©	+
Mn		⊕	⊕	©	©	⊕	©	⊕	©	⊕	⊕		©	⊕	©	+	©
Fe		⊕	⊕	©	⊕	⊕	⊕	⊕	⊕	⊕	⊕	+	⊕	⊕	⊕	+	©
Ga	⊕	⊕	⊕	+	+	©	©	⊕	⊕	⊕	⊕	+	⊕	⊕	⊕	+	⊕
As		©	©	+	+	+	©	©	⊕	⊕	©	+	+	©	+	+	©
Y	↔	↔	⊕	-		-	-	-	↔	↔	-	-	⊕	↔	⊕	-	O
Rh		⊕	⊕	+	⊕	⊕	+	⊕	©	©	⊕	-	©	©	©	+	⊕
In	O	⊕	⊕		+	⊕		↔	↔	↔			⊕	⊕	⊕		
Sb		⊕	+	+	+	⊕	⊕	⊕	⊕	©	⊕	+	©	+	+	+	⊕
La	↔	↔	↔	+	+	©	+	⊕	⊕	⊕	+	©	↔		⊕	⊕	↔
Ce	O	↔	↔	+	+	+	©	+	+	©	+	+	©			⊕	+
Pr	-	↔	↔	+	+	©	+	©	⊕	⊕	+	+	↔	+	⊕		+
Nd	O	↔	↔	+	+	©	+	©	⊕	⊕	+	+	⊕		⊕		+
Pm	-	-	-	+	+	+	+	+	+	+	+	+	+	+	©	+	+
Sm	↔	↔	↔	+	+	+	+	©	⊕	⊕	©	+	⊕		⊕		↔
Eu	-	↔	O	©	+	+	+	+	⊕	©	©	+	⊕	+	⊕	©	
Gd	↔	↔	↔	+	+	©	+	+	⊕	⊕	+	+	⊕	+	⊕	©	↔
Tb	-	↔		+	+	+	+	+	↔	+	+	+	⊕		⊕		+
Dy	-	↔	↔	+	+	©	+	+	↔	©	+	+	⊕	+	⊕	+	+
Ho	O	↔		+	+	+	+	+	↔	©	+	+	⊕	+	⊕	©	+
Er	-	↔	Ø	-	+	+	-		↔		-		⊕	-	©	-	
Tm	-	↔		-	+	+	-		↔		-		⊕		⊕	-	
Yb	↔	↔	⊕	+	+	©	+	+	↔	⊗	+	+	⊕	⊕	⊕	+	
Lu	-		⊕	+	+	+	+	+	↔	+	+	+	⊕	+	⊕	+	+
Au	-	-	©		+								©	+	©		
Tl	-	+	©		©								©	+	©		
Bi	-	↔	©		+						↔		©	©	©		↔
Ac	-	+	+		+								+	+	+		

Fig. 13. Part of table illustrating the prediction of compounds with the composition AB_2O_4. (*Designations:* + formation of compound with composition AB_2O_4 is predicted; − formation of compound with composition AB_2O_4 is not predicted; ⊕ compound with composition AB_2O_4 was synthesized and appropriate information was used in the computer learning process; ↔ compound with composition AB_2O_4 does not exist under normal conditions and this information was used in the computer learning process; © predicted formation of compound with composition AB_2O_4, which was confirmed by experiment; ○ predicted absence of compound with composition AB_2O_4, which is confirmed by experiment; ⊗ predicted formation of compound with composition AB_2O_4, which was not confirmed by experiment; Ø predicted absence of compound with composition AB_2O_4, which was not confirmed by experiment; empty square — indeterminate result).

Shown in Fig. 12 is a part of the table illustrating predictions of the Heusler phases with composition AB_2Cu. All the five checked predictions agreed with the experimental data.

In the process of studies aimed at searching for new magnetic materials we predicted compounds with composition AB_2O_4 (Fig. 13). Of 80 checked predictions, 78 agreed with the new experimental data.

A^+ / B^{3+}	H	Li	Na	K	Cu	Rb	Ag	Cs	Tl	Fr	NH_4
Al			⊕	⊕	+	+	⊕	⊕	⊕	-	⊕
Sc		⊕	⊕	⊕		⊕	⊕	↔	⊕	-	⊕
V			⊕	+	+	©	⊕	©	⊕	+	
Cr			⊕	⊕	+	©	⊕	©	⊕	+	
Fe			⊕	⊕		©	⊕	©		+	©
Ga	+	+	⊕	©	+	©	⊕	©	⊕		
Y		⊕	⊕	⊕		⊕		⊕		-	+
Rh	+	+	⊕	⊕	+	⊕	⊕	⊕	⊕	-	
In				+		+	⊕	+	⊕		⊕
La	+	⊕	⊕	⊕		⊕		⊕			©
Ce		⊕	⊕	⊕		⊕		⊕		-	⊕
Pr		⊕	⊕	⊕		⊕		⊕		-	©
Nd		⊕	⊕	⊕		⊕		⊕		-	+
Pm	-	+	+	+	+	+	+	+	-	-	
Sm	-	⊕	⊕	⊕	+	⊕	+	⊕	-	-	
Eu	-	⊕	⊕	⊕	+	⊕	+	⊕	-	-	
Gd	-	⊕	⊕	⊕	+	⊕	+	⊕	-	-	
Tb	-	⊕	⊕	⊕	+	⊕	+	⊕	-	-	
Dy	-	⊕	⊕	⊕	+	⊕	+	⊕	-	-	
Ho	-	⊕	⊕	⊕	+	⊕	+	⊕	-	-	
Er		⊕	⊕	⊕		⊕		⊕		-	+
Tm	-	⊕	⊕	⊕	+	⊕	+	⊕	-	-	
Yb		⊕	⊕	⊕		⊕		⊕		-	+
Lu	-	⊕	⊕	⊕	+	⊕	+	⊕	-	-	
Bi	-	©	⊕	⊕	+	+	+	+	-	-	
Ac	+	+	+	+	+	+	+	+	-		
Pa	-	+	+	+	+	+	+	+	-	-	
U	-	+	+	+	+	©	+	+	-	-	
Np	-	+	+	+	+	+	+	+	-	-	
Pu	-	+	+	+	+	+	+	+	-	-	
Am	-	+	+	+	+	+	+	+	-	-	

Fig. 14. Part of table illustrating the prediction of compounds with the composition $A_2SO_4{\cdot}B_2(SO_4)_3$. (*Designations:* + formation of compound with composition $A_2SO_4{\cdot}B_2(SO_4)_3$ is predicted; – formation of compound with composition $A_2SO_4{\cdot}B_2(SO_4)_3$ is not predicted; ⊕ compound with composition $A_2SO_4{\cdot}B_2(SO_4)_3$ was synthesized and appropriate information was used in the computer learning process; ↔ compound with composition $A_2SO_4{\cdot}B_2(SO_4)_3$ does not exist under normal conditions and this information was used in the computer learning process; © predicted formation of compound with composition $A_2SO_4{\cdot}B_2(SO_4)_3$ which was confirmed by experiment; empty square — indeterminate result).

Fig. 14 shows results after comparing the predicted double sulphates of monovalent and trivalent elements. All the 14 checked predictions coincided with the new experimental data.

Fig. 15 shows the predictions of new elpasoliths with composition A_2BCF_6. These results were obtained in the process of searching for new substances for laser matrices.

5. Conclusions

We carried out good predictions of qualitative properties of physical–chemical systems: formation of compounds, their crystal structure type, etc. The problem becomes even more complicated if it is necessary to predict some quantitative property (e.g., the melting point, homogeneity range, etc.). The hypothesis of class compactness, based on methods of computer learning presupposes that the different classes locate compactly in the multidimensional feature space and there are no intersections between these classes. However, we found such a set of properties whose space contradict this hypothesis. The application of cluster analysis to the example learning set in combination with the grouping of features according to statistical correlation allows us to decrease the intersections of classes, but only slightly, owing to the selection of the natural (for certain learning set) threshold values of predicted quantitative property. Note that these natural threshold value are less a consequence of the nature of phases and more the set of examples for the computer learning method. These observations are based upon the learning set examples which have been obtained and investigated at present.

Therefore, as a consequence of the above interaction problem, the attempt to predict certain threshold values which are important for technological applications, e.g., boiling point temperatures of helium and nitrogen for superconducting compounds, is justified only from a practical standpoint. The error of this prediction will be high, but it will be possible to predict (with high reliability) those objects which are widely spaced in features' space. A priori identification of these objects by a researcher seems to be a great problem. One possible way to solve this problem is to visualize a two-dimensional projection of points, which corresponds to the objects of the learning set, in combination with cluster analysis of objects and grouping features according to statistical correlation.

As stated above, prediction accuracy of qualitative properties depends strongly on the volume and representativeness of the learning set. Our experience shows that the number of the learning examples must equal 100s or even 1000s in order to have acceptable estimation of quantitative property. However, the solution of tasks of quantitave property prediction is one of the most important problems of our approach.

The future of the approach using DBs and AI methods, is also connected with the development of information-predicting system. That is a very expensive and long procedure. However, such a system allows to cut down the time and expenses involved in the search and development of new materials with predefined properties. It should be noted that this kind of simulation requires only "good" information of DBs.

	A=Na		A=K		A=Rb			A=Cs		A=Tl			
B^{I} C^{III}	K	Tl	Na	Tl	Na	Cs	Tl	K	Tl	Na	K	Rb	Cs
Sc			Ξ		Ξ			Ξ	E	Ξ			*
Ti			Ξ					Ξ	E				*
Mn			Θ		Θ			Θ	E				*
Y	E		Ξ		Ξ			Ξ			Ξ		
Mo	E		Ξ		Ξ								
Rh	E		Ξ										
Pd			Θ		Θ			Θ	E				
Ag	E							Ξ					
In	E	E	Ξ	E	Ξ	E	-	Ξ	Ξ	Ξ	E		
La	⊗		⊗	E				Ξ					
Ce	E	E	E	E		Ξ	E	Ξ	E		E	E	E
Pr	E	E	E	E		E	E	Ξ	E		E		
Nd	E	E	E	E		E	E	Ξ	E		E		
Pm	E	E	E	E		E	E	E	E		E		
Sm	E	E	E	E	Ξ	E	E	Ξ	E		E		
Eu	E	E	E	E	Ξ	E	E	Ξ	E		E		
Gd	E	E	E	E	Ξ	E	E	Ξ	E	E	E		
Tb	E	E	E	E	Ξ	E	E	Ξ	E		E		
Dy	E	E	E	E	Ξ	Ξ	E	Ξ	E		E	E	E
Ho	E	E	E	E	Ξ	E	E	Ξ	E		E		
Er	E	E	E	E	Ξ	E	E	Ξ	E		E		
Tm	E	E	E	E	Ξ	E	E	Ξ	E		E		
Yb	E	E	E	E	Ξ	E	E	Ξ	E		E		
Lu	E	E	E	E	Ξ	E	E	Ξ	E	E	E		
Tl	E		Ξ		Ξ	E		Ξ					
Bi	E	E	E	E	Ξ	E	E	E	E	E	E		
Ac	E	E	E	E	E	E	E	E	E	E	E		*
Pa		E		E	E	E	E		E				*
U	E	E	E	E	E	E	E	E	E	E	E		*
Np	E	E	E	E	E	E	E	E	E	E	E	*	*
Pu	E	E	E	E	E	E	E	E	E	E	E		*
Am	E	E	E	E	E	E	E	E	E	E	E		*
Cm	E	E	E	E	E	E	E	E	E	E	E	*	*
Bk	E	E	E	E	E	E	E	E	E	E	E	*	*
Cf	E	E	E	E	E	E	E	E	E	E	E	*	*

Fig. 15. Prediction of new elpasoliths with composition A_2BCF_6. (*Designations:* E formation of compound with elpasolith crystal structure type is predicted; – formation of compound with elpasolith crystal structure type is not predicted; ∗ formation of compound with composition A_2BCF_6 is not predicted; Ξ compound with elpasolith crystal structure type was synthesized and appropriate information was used in the computer learning process; Θ compound with elpasolith crystal structure type does not exit under normal conditions and this information was used in the computer learning process; ⊗ compound with composition A_2BCF_6 does not exist under normal conditions and this information was used in the computer learning process; empty square — indeterminate result).

Acknowledgements

Partial financial support from RFBR (Grant N.99-07-90040) and EOARD are gratefully acknowledged. I should like to thank my colleagues Prof. Victor P. Gladun and Dr. Neonila D. Vashchenko of Institute of Cybernetics of National Academy of Sciences of Ukraine and Dr. Steven R. Leclair of Wright Laboratories for their help and support.

References

Gladun, V.P., 1994. Processes of Formation of New Knowledge. SD "Pedagog", Sofia, 192.

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

Kiselyova, N.N., Kravchenko, N.V., Petukhov, V.V., 1996. Database system on the properties of ternary inorganic compounds (IBM PC version). Inorganic Materials 32 (5), 567–570.

Kiselyova, N.N., 1997. Application of artificial intelligence methods to inorganic compounds computer design. Perspektivnye Materialy 4, 5–21.

Kiselyova, N.N., 1998. Computational materials design using artificial intelligence methods. J. Alloys and Compounds 279 (1), 8–13.

Kiselyova, N.N., Pokrovskii, B.I., Komissarova, L.N., 1977. Simulation of forming complicated oxides from initial components using cybernetic method of concept formation. Zh. Neorgan. Khimii 22 (4), 883–886.

Savitskii, E.M., Gribulya, V.B., Kiselyova, N.N., 1990. Prediction in Material Science Using Computer. Nauka, Moscow, p. 86.

Savitskii, E.M., Kiselyova, N.N., Pishik, B.N., 1984. Development of data bank on ternary inorganic phase properties. Doklady AN SSSR 279 (3), 627–629.

Yudina, N.V., Petukhov, V.V., Cheremushkin, E.A., 1996. Data bank on acousto-optical, electro-optical, and nonlinear-optical properties of materials. Crystallography Reports 41 (3), 464–468.

Zemskov, V.S., Kiselyova, N.N., 1995. DIAGRAMMA — Database on Phase Diagrams of Semiconductor Systems. Inorganic Materials 31 (9), 1096–1100.