ISSN 2075-1133, Inorganic Materials: Applied Research, 2020, Vol. 11, No. 4, pp. 787–794. © Pleiades Publishing, Ltd., 2020.
Russian Text © The Author(s), 2019, published in Materialovedenie, 2019, No. 9, pp. 41–48.

Computer-Aided Design of Compounds with Crystal Structure of Melilites

N. N. Kiselyova[a], *, V. A. Dudarev[a], V. V. Ryazanov[b], O. V. Sen'ko[b], and A. A. Dokukin[b]

[a] *Baikov Institute of Metallurgy and Materials Science, Russian Academy of Sciences (IMET RAS), Moscow, 119334 Russia*
[b] *Federal Research Center Informatics and Management, Russian Academy of Sciences, Moscow, 119333 Russia*
**e-mail: kis@imet.ac.ru*

Received November 1, 2018; revised February 6, 2019; accepted April 16, 2019

Abstract—Compounds of compositions $A_2^{+2}B_2^{+3}C^{+4}O_7$ and $A^{+2}B_2^{+2}C_2^{+4}O_7$ that are not yet obtained (A and B are cations of different elements; C is Si or Ge) with a melilite-type crystal structure are predicted and their crystal lattice parameters are evaluated. Predicting is based only on data on the properties of elements and simple oxides. The mean accuracy of predicting is at least 85%. The calculations are performed using scikit-learn system programs and an information analytical system based on machine learning approaches.

Keywords: melilite, database, design of inorganic compounds, machine learning
DOI: 10.1134/S2075113320040188

INTRODUCTION

Compounds of composition $A_2^{+2}B_2^{+3}C^{+4}O_7$ and $A^{+2}B_2^{+2}C_2^{+4}O_7$ (A and B are cations of different elements; C is Si or Ge) with a crystal structure such as melilite (space group is $P4\,(-)\,2_1m$, $Z = 2$) are of interest for the search for new nonlinear optical [1], piezoelectric [2], and ferroelectric [3] materials, multiferroics [4], etc. [5, 6].

The databases Phases [7] and Bandgap [8] store information on 36 melilites of composition $A_2^{+2}B_2^{+3}C^{+4}O_7$ and on 38 melilites of composition $A^{+2}B_2^{+2}C_2^{+4}O_7$. Most of these compounds are formed in systems with SiO_2 and GeO_2.

The purpose of this work is to predict new compounds with a crystal structure such as melilite in systems of $AO–B_2O_3–SiO_2$ (GeO_2) and $AO–BO–SiO_2$ (GeO_2).

FORMULATION OF THE PROBLEM

We found several articles [9–11] the authors of which tried to find the relationship between the type of crystal structure of the compounds and the ionic radii (according to Shannon) of the components. The stability diagram of melilites of composition $A_2^{+2}B_2^{+3}C^{+4}O_7$ in the space of ionic radii of cations A^{+2} and B^{+3} is proposed in [9]. We analyzed the location of points in the coordinates of the ionic radii of the cations A^{+2} and B^{+3}; these points correspond to compounds of composition $A_2^{+2}B_2^{+3}C^{+4}O_7$ with a melilite structure, as well as compounds with a different crystal structure and systems of $AO–BO–SiO_2$ (GeO_2) without the formation of compounds with the above composition. The analysis showed that there is a significant intersection of the regions corresponding to the compounds $A_2^{+2}B_2^{+3}C^{+4}O_7$ and systems without the formation of compounds with this composition. Similar attempts to use the ionic radii of components A and C to search for regions of existence of melilites with compositions $A_2^{+2}B_2^{+3}C^{+4}O_7$ [10] and $A^{+2}B_2^{+2}C_2^{+4}O_7$ [11] also led to a significant intersection of regions corresponding to different classes of substances.

In order to clarify the boundaries of different classes of compounds with compositions $A_2^{+2}B_2^{+3}C^{+4}O_7$ and $A^{+2}B_2^{+2}C_2^{+4}O_7$ and to increase the accuracy of predicting the possibility of formation and type of crystal structure of compounds, we tried to expand the set of properties of components A, B, and C. To search for multidimensional criteria, we propose to use machine learning approaches.

EXPERIMENTAL

Calculation Methods

To solve the problem, we developed an information analytical system (IAS) [12] which combines databases on the properties of inorganic substances and materials and an information analysis subsystem based

on machine learning (precedent based pattern recognition). Procedures for using IAS to search for complex regularities in chemical information and predicting new inorganic compounds and estimate their properties included the following steps.

(1) Selection of examples for computer analysis.

(2) Selection of the initial set of component properties (elements and simple oxides) for the formation of the desired criteria.

(3) Finding the properties of components and simple algebraic functions of these properties; these functions would ensure the best separation of various classes of compounds of the above composition.

(4) Computer analysis of the selected information with a further selection of criteria separating the classes of substances in the best way.

(5) The use of the found multidimensional criteria for predicting compounds that are not yet obtained and estimating their properties.

Selection of Examples for Computer Analysis

The main source of initial information for computer analysis is the database Phases [7], which is part of the IAS. One of the problems that significantly reduces the accuracy of prediction is extremely scarce information on compounds of predicting composition with crystal structures other than melilite, as well as on the systems of $AO-B_2O_3-SiO_2$ (GeO_2) and $AO-BO-SiO_2$ (GeO_2), in which compounds of the above compositions are not formed.

Selection of the Initial Set of Component Properties for the Formation of the Desired Criteria

Information on the compounds is represented in computer memory in the form of a matrix the rows of which include a set of parameter values of the components of a particular compound, indicating to which class this compound belongs. The physical and chemical nature of the studied substances is taken into account when choosing the initial properties of the components. Information on the properties of elements is obtained from the database Elements (http://phases.imet-db.ru/elements) that we developed, and information on oxides is obtained from the database on the properties of simple oxides included in the IAS. The result of the first two stages is a sample formed for subsequent computer analysis (learning sample).

Finding the Properties of the Components Most Important for Classification and Simple Algebraic Functions of These Properties

A special program [13] included in the IAS is used to select the properties of the components that are most important for the separation of different classes of substances. Both the initial properties and automatically generated algebraic functions of these properties are evaluated in the analysis of the learning sample, which greatly simplifies the subsequent formation of the desired criteria. However, despite the fact that the generation of functions is carried out using a set of elementary algebraic operations on the values of the parameters of components, which are the same in physical sense and dimension, the number of parameters generated in this way often exceeds several hundred and even thousands. In this regard, the selection of only algebraic functions which are the most important for classification so that they can be included in the desired criterion significantly accelerates its formation and often helps to increase the accuracy of predicting. The result of the programs is to find the parameters of the components that most separated the given classes.

Computer Analysis of Selected Information

A set of 15 programs of precedent based pattern recognition included in IAS [12, 14] is used to search for criteria for the formation of compounds and criteria for predicting the type of their crystal structure under normal conditions. The cross-validation procedure, which is described in detail in [14], is used to assess the accuracy of predicting the formed criteria. As a result, the best pattern recognition algorithms are selected for solving this problem. The collective decision-making procedure based on special programs [14] included in the IAS is used to formulate a generalized criterion using the advantages of various algorithms. Examination recognition of information about a given number of substances, the data on which are randomly selected from learning samples and are not used in machine learning (at the final stage of predicting, these control examples are returned to the sample for analysis), is used to estimate the accuracy of the criteria obtained using these programs.

Predicting of the Formation of Compounds Not Yet Obtained

Only data on the properties of the components are used in predicting new compounds. All predictions are carried out for the case of atmospheric pressure and room temperature. The procedure of predicting and generating a table of predictions is carried out automatically in the IAS. The user specifies only symbol sets of components.

All these procedures of machine learning and predicting are carried out separately for compounds of compositions $A_2^{+2}B_2^{+3}C^{+4}O_7$ and $A^{+2}B_2^{+2}C_2^{+4}O_7$.

CALCULATIONS

Compounds of Composition $A_2^{+2}B_2^{+3}C^{+4}O_7$

After an expert analysis, information on 38 compounds of composition $A_2^{+2}B_2^{+3}C^{+4}O_7$ with the melilite structure, on 10 compounds with a crystal structure different from melilite under ambient conditions, and on 26 systems of $AO-B_2O_3-CO_2$ in which the compounds of composition $A_2B_2CO_7$ are not formed are included in the sample for computer analysis.

Compounds of Composition $A^{+2}B_2^{+2}C_2^{+4}O_7$

Information on 36 compounds of composition $A^{+2}B_2^{+2}C_2^{+4}O_7$ with a melilite structure, on 10 compounds with a crystal structure different from melilite under ambient conditions, and on 14 systems of $AO-BO-CO_2$ in which the compounds of composition $AB_2C_2O_7$ are not formed are included in the sample for computer analysis.

In both tasks, the initial set of parameters of components includes the properties of chemical elements A, B, C, and O: pseudopotential radius (according to Zunger), ionic radius (according to Shannon), distances to core and valence electrons (according to Schubert), ionization energies of the first, second, and third electrons (E5–E7), the numbers (according to Mendeleev and Pettifor (M1–M11)), quantum number (A5), electronegativity (according to Pauling), Miedema's chemical potential, melting point and boiling point, standard entropy, atomization enthalpy, molar heat capacity, thermal conductivity, etc., as well as the thermal parameters of the simple oxides AO, B_2O_3, and CO_2 (melting point (decomposition), standard entropy, heat of formation, isobaric heat capacity, and isobaric formation potential). There are a total of 105 values for each system of A–B–C–O.

RESULTS AND DISCUSSION

Predicting of Melilites of Composition $A_2^{+2}B_2^{+3}C^{+4}O_7$

The solution to the task of predicting the possibility of the formation of compounds $A_2^{+2}B_2^{+3}C^{+4}O_7$ allowed us to conclude that the ratio of Mendeleev's numbers [15] of "t-d start right" of element A and "H d-t start right" of element B (one can see the parameter values in database Elements http://phases.imet-db.ru/elements) (M9(A)/M4(B)) and the ratio of the ionization energy of the second electron of element B to the ionization energy of the third electron of element C (E7(B)/E5(C)) turned out to be the most separating classes of compound formation and lack of compound. However, the estimation of predicting accuracy showed that using only these informative parameters did not allow achieving higher accuracy than when 105 initial parameters of elements and simple oxides are included in the separating multidimensional criterion. In the latter case, the use of a group of neural network learning algorithms, methods of *k*-nearest neighbors and support vector machine, and the convex stabilization method for the collective decision-making procedure made it possible to achieve predicting with an accuracy of 93.3%.

Next, we predicted compounds of composition $A_2^{+2}B_2^{+3}C^{+4}O_7$ with the structure of melilite under normal conditions. Analysis of the learning sample showed that functions such as the ratio of Mendeleev's numbers "d-t start right" of element A and "t-d start right" of element C (M11(A)/M9(C)) and the difference between the ionization energy of the first electron of element B and the ionization energy of the third electron of element C (E5(B)–E7(C)) have the highest separation ability. In this task, adding the found informative parameters (M11(A)/M9(C) and E5(B)–E7(C)) to the initial set of properties of elements and simple oxides made it possible to achieve predicting with an accuracy of 100% when applying a group of algorithms based on the methods of a linear machine, the formation of logical regularities, support vector machine, *k*-nearest neighbors, and training a neural network, as well as the Bayesian approach for making a collective decision.

The formed criteria were used to predict the possibility of formation and type of crystal structure under normal conditions of compounds that are not yet synthesized. Tables 1 and 2 contain part of the results of comparison of predictions obtained in solving both classification tasks. The following notation is accepted: the prediction of compounds with a crystal structure such as melilite (M); the prediction of compounds with a crystalline structure other than melilite (A); the prediction of the absence of a compound with composition $A_2^{+2}B_2^{+3}Si^{+4}O_7$ in system $AO-B_2O_3-CO_2$ (N); empty cells are discrepancies in the results of comparison of the predictions of compound formation and the type of crystal structure. The symbol # indicates previously studied systems, information about which was used in the formation of multidimensional criteria.

Predicting of Melilites of Composition $A^{+2}B_2^{+2}C_2^{+4}O_7$

The search for informative features that would allow the best separation of systems $AO-BO-CO_2$ with and without formation of the compounds of composition $AB_2C_2O_7$ under normal conditions allowed us to find two functions; they are the difference between the ionization energy of the first electron of element A and the ionization energy of the third electron of element C (E5(A)–E7(C)) and the ratio of the Mendeleev–Pettifor number of element A and the quantum number of element B (M5(A)/A5(B)). Both classes are relatively well separated. Only two substances fall into other regions (a system $CaO-NiO-SiO_2$ and a compound $Pb_2ZnSi_2O_7$).

Table 1. Predicting of the type of crystal structure of compounds with composition $A_2^{+2}B_2^{+3}Si^{+4}O_7$

B	A						
	Be	Mg	Ca	Cu	Zn	Sr	Ba
B		M	#M	N	N	M	M
Al		#N	#M	#N	N	#M	#N
Cr	N	#N		N	N	M	M
Fe	N	N	#N	N	N	#M	
Ga		N	#M	N	N	M	#N
Y	#M	A	M	N		M	A
La	M	A	A	N		A	A
Ce	M	A	A	N		A	A
Pr	M	A	M	N		A	A
Nd	#M	A	A	N		A	A
Pm	M	A	M	N		M	A
Sm	#M	A	M	N		M	A
Eu	#M	M	M	N	N	M	A
Gd	#M	M	M	N		M	A
Tb	#M	A	M	N		M	A
Dy	#M	A	M	N	N	M	
Ho	#M	M	M	N	N	M	
Er	#M	M	M	N	N	M	
Tm	#M	M	M	N	N	M	
Yb	#M		M	N	N	M	
Lu	#M	M	M	N	N	M	
Bi		N	M	N	N	M	

Table 2. Predicting of the type of crystal structure of compounds with composition $A_2^{+2}B_2^{+3}Ge^{+4}O_7$

B	A						
	Be	Mg	Ca	Cu	Zn	Sr	Ba
B	M	M	M	N	N	M	M
Al		M	#M	N	N	#M	M
Cr		N		N	N	M	M
Fe	N	N	M	N	N	#M	#M
Ga	M	M	#M	N	N	#M	#M
Y	#M	A	A	A	A	A	M
La	#M	A	A	A	A	M	M
Ce	M	A	A	A	A	A	A
Pr	#M	A	A	A	A	A	A
Nd	#M	A	A	A	A	M	M
Pm	M	A	A	A	A	M	M
Sm	#M	A	M	A	A	M	M
Eu	#M	M	A	A	A	A	A
Gd	#M	A	A	A	A	A	A
Tb	#M	A	A	A	A	A	A
Dy	#M	A	A	A	A	M	M
Ho	#M	M	A	A	A	M	M
Er	#M	M	A	A	A	M	M
Tm	#M	M	A	N		M	M
Yb	#M	M	M	N		M	M
Lu	#M	M	M	N	A	M	M
Bi	M		M	#N	N	M	M

The most informative parameters for separating melilites from compounds with other crystal structures and systems without the formation of a compound with composition $A^{+2}B_2^{+2}C_2^{+4}O_7$ are the Mendeleev number "d-t start right" of element A (M11(A)), the difference between the ionization energy of the second electron of element A and the ionization energy of the third electron of element C (E6(A)–E7(C)), and the ratio of the ionization energy of the first electron of element A to the ionization energy of the third electron of element C (E5(A)/E7(C)). However, the analysis shows that the use of the Mendeleev number M11(A) and the above informative functions does not make it possible to separate the stability region of melilites from the regions of other substances. The compounds $Ba_2ZnSi_2O_7$ and $Ca_2ZnGe_2O_7$ and a system $CaO–NiO–SiO_2$ without the formation of compound $AB_2C_2O_7$ fall into the melilite region. It should be noted that the data on the crystal structure of $Ba_2ZnSi_2O_7$ are contradictory. A publication in 1974 [16] classifies it as melilites, and later studies [17] indicate a monoclinic structure with a space group C2/c. The high-temperature crystal modification of $Ca_2ZnGe_2O_7$ also has a melilite structure [18]; however, it transforms into monoclinic polymorphs with a decrease in temperature [18].

Tables 3 and 4 contain a part of the results of comparison of predictions obtained in solving two classification tasks (the formation of compounds and the type of crystal structure). A group of algorithms including Fisher's linear discriminant, voting according to logical regularities, the method of support vector machine, and the complex committee method (averaging) for making a collective decision is used in predicting. The resulting classification rule includes only the properties of the elements. This rule makes it possible to achieve predicting accuracy of 85%. The notation is the same as in Tables 1 and 2.

Predicting of Crystal Lattice Parameters of Melilites

Predicting of the crystal lattice parameters of compounds is of great interest for both chemical research and materials science investigation. Machine learning methods are widely used to solve this problem. For example, in [19–22], the crystal lattice parameters of

Table 3. Predicting of the type of crystal structure of compounds with composition $A^{+2}B_2^{+2}Si_2^{+4}O_7$

B	A				
	Ca	Sr	Ba	Eu	Pb
Be	#M	M	M	M	#N
Mg	#M	#M	#A	#M	#N
Mn	#M	#M	#M	#M	#M
Fe	#M	#M	#M	M	#M
Co	#M	#M	#M	M	M
Ni	#N	M		M	
Cu	M	#M	#A	M	M
Zn	#M	#M	#A	M	#M
Cd	#M	#M	#M	M	#N

Table 4. Predicting of the type of crystal structure of compounds with composition $A^{+2}B_2^{+2}Ge_2^{+4}O_7$

B	A				
	Ca	Sr	Ba	Eu	Pb
Be	#N	#N	N	N	N
Mg	#M	#M	#M	M	N
Mn	M	#M	#M	M	M
Fe	M	#M	#M	M	M
Co	M	#M	#M	M	M
Ni	M	M	M	M	M
Cu	M	M	#M	M	M
Zn	#A	#M	#M	M	M
Cd	M	M	#M	M	M

orthorhombic perovskites with composition ABO_3 not yet obtained were predicted using the methods of neural network training and support vector machine. Using the neural network training and regression on support vector machine, it was possible to predict the crystal lattice parameters of cubic and monoclinic perovskites with composition ABX_3 (X is halogen or oxygen) [23]. The same methods and random forest learning were used to predic the crystal lattice parameters of cubic perovskites of composition $A_2^{+2}BCO_6$ [24, 25] and apatites [26–28]. The lattice parameters and band gap were predicted for compounds of composition ABX_2 with chalcopyrite structure using neural network training and various statistical methods (discriminate analysis, principal component analysis, etc.) [29].

The crystal lattice parameters of the predicted melilites were evaluated in several stages.

1. The selection of the initial set of properties of components (elements and simple oxides) for the formation of learning samples.

Information on the compounds was presented in the form of a matrix the rows of which included a set of values of the component parameters of a particular compound with the parameter values "*a*" or "*c*" of the crystal lattice. Taking into account two target parameters ("*a*" and "*c*") and two types of substance compositions ($A_2^{+2}B_2^{+3}C^{+4}O_7$ and $A^{+2}B_2^{+2}C_2^{+4}O_7$), we have four matrices for learning and, accordingly, four tasks on predicting (depending on the type of parameters and composition of the compounds). The properties of the elements and simple oxides were chosen as the initial properties of the components. We also added the most informative algebraic functions of the initial parameters of the components found as a result of solving the prediction tasks of the type of crystal structure. As a result, four learning samples were generated for subsequent computer analysis. For melilites with composition $A_2^{+2}B_2^{+3}C^{+4}O_7$, the learning samples included 37 examples, and for $A^{+2}B_2^{+2}C_2^{+4}O_7$, the learning samples included 28 examples.

2. Computer analysis of the selected information and further selection of the best algorithms.

Table 5. Predicting results of the crystal lattice parameters of melilites with composition $A^{+2}B_2^{+2}C_2^{+4}O_7$ not yet studied

Composition	*a*, Å	*c*, Å	Composition	*a*, Å	*c*, Å
$Sr_2BeSi_2O_7$	7.596	5.138	$Pb_2FeGe_2O_7$	8.443	5.119
$Ba_2BeSi_2O_7$	7.777	5.328	$Ca_2CoGe_2O_7$	7.947	5.161
$Eu_2BeSi_2O_7$	7.592	5.145	$Eu_2CoGe_2O_7$	8.188	5.339
$Eu_2FeSi_2O_7$	8.102	5.150	$Pb_2CoGe_2O_7$	8.350	5.144
$Eu_2CoSi_2O_7$	8.010	5.176	$Ca_2NiGe_2O_7$	7.950	5.139
$Pb_2CoSi_2O_7$	8.197	4.989	$Sr_2NiGe_2O_7$	8.164	5.300
$Sr_2NiSi_2O_7$	8.016	5.146	$Ba_2NiGe_2O_7$	8.383	5.502
$Eu_2NiSi_2O_7$	8.012	5.154	$Eu_2NiGe_2O_7$	8.191	5.317
$Ca_2CuSi_2O_7$	7.918	4.930	$Pb_2NiGe_2O_7$	8.353	5.122
$Eu_2CuSi_2O_7$	8.091	5.087	$Ca_2CuGe_2O_7$	8.028	5.072
$Pb_2CuSi_2O_7$	8.278	4.900	$Sr_2CuGe_2O_7$	8.242	5.234
$Eu_2ZnSi_2O_7$	7.990	5.182	$Eu_2CuGe_2O_7$	8.269	5.250
$Eu_2CdSi_2O_7$	8.260	5.191	$Pb_2CuGe_2O_7$	8.431	5.056
$Eu_2MgGe_2O_7$	8.200	5.326	$Eu_2ZnGe_2O_7$	8.169	5.345
$Ca_2MnGe_2O_7$	8.069	5.157	$Pb_2ZnGe_2O_7$	8.331	5.151
$Eu_2MnGe_2O_7$	8.310	5.335	$Ca_2CdGe_2O_7$	8.197	5.176
$Pb_2MnGe_2O_7$	8.472	5.140	$Sr_2CdGe_2O_7$	8.411	5.337
$Ca_2FeGe_2O_7$	8.040	5.135	$Eu_2CdGe_2O_7$	8.438	5.354
$Eu_2FeGe_2O_7$	8.281	5.313	$Pb_2CdGe_2O_7$	8.600	5.160

Table 6. Predicting results of the crystal lattice parameters of melilites with composition $A_2^{+2}B_2^{+3}C^{+4}O_7$ not yet studied

Composition	a, Å	c, Å	Composition	a, Å	c, Å
$Mg_2B_2SiO_7$	6.737	4.719	$Ca_2B_2GeO_7$	7.220	4.876
$Sr_2B_2SiO_7$	7.294	5.028	$Sr_2B_2GeO_7$	7.382	5.099
$Ba_2B_2SiO_7$	7.696	5.230	$Ba_2B_2GeO_7$	7.778	5.312
$Sr_2Cr_2SiO_7$	8.007	5.146	$Ba_2Al_2GeO_7$	8.316	5.552
$Ba_2Cr_2SiO_7$	8.409	5.348	$Ca_2Cr_2GeO_7$	7.933	4.994
$Sr_2Fe_2SiO_7$	8.151	5.270	$Sr_2Cr_2GeO_7$	8.096	5.216
$Sr_2Ga_2SiO_7$	7.956	5.337	$Ba_2Cr_2GeO_7$	8.496	5.429
$Ca_2Y_2SiO_7$	8.231	5.076	$Ca_2Fe_2GeO_7$	8.078	5.118
$Sr_2Y_2SiO_7$	8.400	5.287	$Be_2Ga_2GeO_7$	6.951	4.842
$Be_2La_2SiO_7$	7.574	4.921	$Mg_2Ga_2GeO_7$	7.501	5.075
$Be_2Ce_2SiO_7$	7.541	4.916	$Be_2Ce_2GeO_7$	7.651	4.952
$Be_2Pr_2SiO_7$	7.502	4.883	$Be_2Pm_2GeO_7$	7.577	4.915
$Ca_2Pr_2SiO_7$	8.447	5.203	$Ca_2Sm_2GeO_7$	8.454	5.200
$Be_2Pm_2SiO_7$	7.468	4.879	$Sr_2Sm_2GeO_7$	8.616	5.423
$Ca_2Pm_2SiO_7$	8.414	5.199	$Mg_2Eu_2GeO_7$	8.040	5.076
$Sr_2Pm_2SiO_7$	8.583	5.410	$Ca_2Eu_2GeO_7$	8.421	5.186
$Mg_2Sm_2SiO_7$	7.970	5.043	$Sr_2Eu_2GeO_7$	8.584	5.408
$Ca_2Sm_2SiO_7$	8.358	5.142	$Ca_2Gd_2GeO_7$	8.408	5.184
$Sr_2Sm_2SiO_7$	8.527	5.352	$Sr_2Gd_2GeO_7$	8.571	5.406
$Mg_2Eu_2SiO_7$	7.938	5.029	$Ca_2Tb_2GeO_7$	8.377	5.151
$Ca_2Eu_2SiO_7$	8.326	5.127	$Sr_2Tb_2GeO_7$	8.540	5.373
$Sr_2Eu_2SiO_7$	8.495	5.338	$Mg_2Dy_2GeO_7$	7.974	5.039
$Mg_2Gd_2SiO_7$	7.925	5.027	$Ca_2Dy_2GeO_7$	8.355	5.149
$Ca_2Gd_2SiO_7$	8.313	5.125	$Sr_2Dy_2GeO_7$	8.517	5.371
$Sr_2Gd_2SiO_7$	8.482	5.336	$Mg_2Ho_2GeO_7$	7.947	5.024
$Ca_2Tb_2SiO_7$	8.282	5.092	$Ca_2Ho_2GeO_7$	8.328	5.134
$Sr_2Tb_2SiO_7$	8.451	5.303	$Sr_2Ho_2GeO_7$	8.491	5.357
$Ca_2Dy_2SiO_7$	8.259	5.090	$Mg_2Er_2GeO_7$	7.927	5.014
$Sr_2Dy_2SiO_7$	8.428	5.301	$Ca_2Er_2GeO_7$	8.308	5.124
$Mg_2Ho_2SiO_7$	7.845	4.977	$Sr_2Er_2GeO_7$	8.471	5.347
$Ca_2Ho_2SiO_7$	8.233	5.076	$Mg_2Tm_2GeO_7$	7.908	4.993
$Sr_2Ho_2SiO_7$	8.402	5.286	$Ca_2Tm_2GeO_7$	8.289	5.103
$Mg_2Er_2SiO_7$	7.825	4.967	$Sr_2Tm_2GeO_7$	8.451	5.325
$Ca_2Er_2SiO_7$	8.213	5.066	$Ba_2Tm_2GeO_7$	8.846	5.538
$Sr_2Er_2SiO_7$	8.382	5.276	$Mg_2Yb_2GeO_7$	7.887	4.989
$Mg_2Tm_2SiO_7$	7.805	4.946	$Ca_2Yb_2GeO_7$	8.268	5.099
$Ca_2Tm_2SiO_7$	8.193	5.044	$Sr_2Yb_2GeO_7$	8.430	5.321
$Sr_2Tm_2SiO_7$	8.362	5.255	$Ba_2Yb_2GeO_7$	8.825	5.534
$Ca_2Yb_2SiO_7$	8.172	5.040	$Mg_2Lu_2GeO_7$	7.866	4.981
$Sr_2Yb_2SiO_7$	8.341	5.251	$Ca_2Lu_2GeO_7$	8.247	5.091
$Mg_2Lu_2SiO_7$	7.764	4.934	$Sr_2Lu_2GeO_7$	8.410	5.314
$Ca_2Lu_2SiO_7$	8.152	5.033	$Ba_2Lu_2GeO_7$	8.805	5.527
$Sr_2Lu_2SiO_7$	8.321	5.243	$Mg_2Bi_2GeO_7$	7.809	5.315
$Ca_2Bi_2SiO_7$	8.094	5.366	$Ca_2Bi_2GeO_7$	8.190	5.425
$Sr_2Bi_2SiO_7$	8.264	5.577	$Sr_2Bi_2GeO_7$	8.352	5.647
$Mg_2B_2GeO_7$	6.839	4.767	$Ba_2Bi_2GeO_7$	8.747	5.860

The complex of linear regression methods included in the scikit-learn package [30] was used to search for methods that make it possible to predict the value of the crystal lattice parameters under normal conditions. Determination coefficient (R^2) [31] was used to estimate the quality of functions formed as a result of the work of the corresponding programs.

The classical least squares method (linear_model.LinearRegression algorithm) and the Theil–Sen method of robust linear smoothing of a set of points [32, 33] (linear_model.TheilSenRegressor algorithm) showed very good results ($R^2 > 0.99$). The advantage of the Theil–Sen method is insensitivity to outliers in the data, which makes it possible to use it in the case of noisy or poor-quality initial information. The algorithms which showed the best results were selected to solve the problem of evaluating the parameters of the crystal lattice of melilites. The predictions of the crystal lattice parameters of melilites with compositions $A^{+2}B_2^{+2}C_2^{+4}O_7$ (Table 5) and $A_2^{+2}B_2^{+3}C^{+4}O_7$ (Table 6) were obtained using averaging of the values found by these two methods.

CONCLUSIONS

Analysis of the database information on the properties of inorganic compounds using machine learning approaches made it possible to predict compounds that are not yet synthesized with compositions $A_2^{+2}B_2^{+3}C^{+4}O_7$ and $A^{+2}B_2^{+2}C_2^{+4}O_7$ (A and B are cations of different elements; C is Si or Ge) with a crystal structure such as melilite. The use of machine learning methods made it possible to accurately evaluate the crystal lattice parameters of melilites. Predicting was based only on data on the properties of chemical elements and simple oxides.

Predicting compounds with the acentric crystalline structure of melilite are promising for the search for new nonlinear optical, piezoelectric, and ferroelectric materials.

FUNDING

This work was supported in part by the Russian Foundation for Basic Research, project nos. 17-07-01362 and 18-07-00080. The study was carried out as part of the state assignment (project nos. 007-00129-18-00 and 0063-2020-0003).

REFERENCES

1. Becker, P., Bohaty L., Liebertz J., et al., Non-centrosymmetric tetragonal $Sr_2ZnGe_2O_7$—a novel melilite-type nonlinear-laser crystal offering $\chi^{(2)}$-, $\chi^{(3)}$-, and cascaded $\chi^{(3)} \leftrightarrow \chi^{(2)}$-interactions, *Laser Phys. Lett.*, 2010, vol. 7, no. 5, pp. 367–377.
2. Shen, C., Zhang, S., Cao, W., et al., Thermal and electromechanical properties of melilite-type piezoelectric single crystals, *J. Appl. Phys.*, 2015, vol. 117, no. 6, art. ID 064106.
3. Zou, Z.-Y., Lan, X.-K., Lu, W.-Z., et al., Novel high Curie temperature $Ba_2ZnSi_2O_7$ ferroelectrics with low-permittivity microwave dielectric properties, *Ceram. Int.*, 2016, vol. 42, no. 14, pp. 16387–16391.
4. Hutanu, V., Sazonov, A., Meven, M., et al., Determination of the magnetic order and the crystal symmetry in the multiferroic ground state of $Ba_2CoGe_2O_7$, *Phys. Rev. B*, 2012, vol. 86, no. 10, art. ID 104401.
5. Shen, C., Wang, D., Xu, H., et al., Bulk crystal growth and thermal, spectroscopic and laser properties of disordered melilite Nd: $Ca_2Ga_2SiO_7$ single crystal, *J. Alloys Compd.*, 2017, vol. 727, pp. 8–13.
6. Luo, Y. and Xia, Z., Effect of partial nitridation on the structure and luminescence properties of melilite-type $Ca_2Al_2SiO_7$:Eu^{2+} phosphor, *Opt. Mater.*, 2014, vol. 36, no. 11, pp. 1874–1878.
7. Kiseleva, N.N., Murat, D., Stolyarenko, A., et al., Database on "phases" ternary inorganic compound properties in Internet, *Inf. Russ. Resur.*, 2006, no. 4, pp. 21–23.
8. Kiselyova, N.N., Dudarev, V.A., and Korzhuyev, M.A., Database on the bandgap of inorganic substances and materials, *Inorg. Mater.: Appl. Res.*, 2016, vol. 7, no. 1, pp. 34–39.
9. Ochi, Y., Morikawa, H., Minato, I., and Marumo, F., Preparation and magnetic property of new rare earth compounds $R_2GeBe_2O_7$ (R = La, Pr, Sm, Gd, Dy, Er) and $Y_2GeBe_2O_7$, *Mater. Res. Bull.*, 1982, vol. 17, no. 7, pp. 911–916.
10. Mill', B.V. and Baibakova, G.D., New compounds with melilite structure: $Ln_2SiBe_2O_7$ (Ln = Nd–Eu, Tb, Ho–Lu) and $Ln_2GeBe_2O_7$ (Ln = Nd, Eu, Tb, Ho, Tm–Lu), *Zh. Neorg. Khim.*, 1990, vol. 35, no. 3, pp. 604–607.
11. Mill', B.V. and Baibakova, G.D., $Ba_2CdGe_2O_7$ is a terminal element of okermanite series, *Zh. Neorg. Khim.*, 1990, vol. 35, no. 3, pp. 789–791.
12. Kiselyova, N.N., Stolyarenko, A.V., Ryazanov, V.V., et al., A system for computer-assisted design of inorganic compounds based on computer training, *Pattern Recognit. Image Anal.*, 2011, vol. 21, no. 1, pp. 88–94.
13. Senko, O.V., An optimal ensemble of predictors in convex correcting procedures, *Pattern Recognit. Image Anal.*, 2009, vol. 19, no. 3, pp. 465–468.
14. Zhuravlev, Yu.I., Ryazanov, V.V., and Sen'ko, O.V., *Raspoznavanie. Matematicheskie metody. Programmnaya sistema. Prakticheskie primeneniya* (Recognition. Mathematical Methods. Program System. Practical Use), Moscow: Fazis, 2006.
15. Pettifor, D.G., A chemical scale for crystal-structure maps, *Solid State Commun.*, 1984, vol. 51, no. 1, pp. 31–34.
16. Sirazhiddinov, N.A., Mirababaeva, N.N., Grebenshchikov, R.G., and Stroganov, E.V., Isomorphism of barium germanosilicates in the Ba_2SiO_4,Ba_2GeO_4‖ $MSiO_3$,$MGeO_4$ quaternary reciprocal system, *Russ. J. Inorg. Chem.*, 1974, vol. 19, no. 3, pp. 817–818.
17. Kaiser, J.W. and Jeitschko, W., Crystal structure of the new barium zinc silicate $Ba_2ZnSi_2O_7$, *Z. Kristallogr.*, 2002, vol. 217, no. 1, pp. 25–26.

18. Armbruster, T., Rothlisberger, F., and Seifert, F., Layer topology, stacking variation, and site distortion in melilite-related compounds in the system CaO–ZnO–GeO_2–SiO_2, *Am. Miner.*, 1990, vol. 75, pp. 847–858.

19. Aleksovska, S., Dimitrovska, S., and Kuzmanovski, I., Crystal structure prediction in orthorhombic ABO_3 perovskites by multiple linear regression and artificial neural networks, *Acta Chim. Sloven.*, 2007, vol. 54, no. 3, pp. 574–582.

20. Javed, S.G., Khan, A., Majid, A., et al., Lattice constant prediction of orthorhombic ABO_3 perovskites using support vector machines, *Comput. Mater. Sci.*, 2007, vol. 39, no. 3, pp. 627–634.

21. Khan, A. and Javed, S.G., Predicting regularities in lattice constants of $GdFeO_3$-type perovskites, *Acta Crystallogr.*, 2008, vol. 64, no. 1, pp. 120–122.

22. Li, C., Thing, Y., Zeng, Y., et al., Prediction of lattice constant in perovskites of $GdFeO_3$ structure, *J. Phys. Chem. Solids*, 2003, vol. 64, no. 11, pp. 2147–2156.

23. Majid, A., Khan, A., Javed, G., and Mirza, A.M., Lattice constant prediction of cubic and monoclinic perovskites using neural networks and support vector regression, *Comput. Mater. Sci.*, 2010, vol. 50, no. 2, pp. 363–372.

24. Dimitrovska, S., Aleksovska, S., and Kuzmanovski, I., Prediction of the unit cell edge length of cubic $A_2^{2+}BB'O_6$ perovskites by multiple linear regression and artificial neural networks, *Central Eur. J. Chem.*, 2005, vol. 3, no. 1, pp. 198–215.

25. Majid, A., Khan, A., and Choi, T.-S., Predicting lattice constant of complex cubic perovskites using computational intelligence, *Comput. Mater. Sci.*, 2011, vol. 50, no. 6, pp. 1879–1888.

26. Kockan, U. and Evis, Z., Prediction of hexagonal lattice parameters of various apatites by artificial neural network, *J. Appl. Cryst.*, 2010, vol. 43, no. 4, pp. 769–779.

27. Legrain, F., Carrete, J., van Roekeghem, A., et al., Materials screening for the discovery of new half-Heuslers: machine learning versus ab initio methods, *J. Phys. Chem.*, 2018, vol. 122, no. 2, pp. 625–632.

28. Oliynyk, A.O., Adutwum, L.A., Rudyk, B.W., et al., Disentangling structural confusion through machine learning: structure prediction and polymorphism of equiatomic ternary phases *ABC, J. Am. Chem. Soc.*, 2017, vol. 139, no. 49, pp. 17870–17881.

29. Zeng, Y., Chua, S.J., and Wu, P., On the prediction of ternary semiconductor properties by artificial intelligence methods, *Chem. Mater.*, 2002, vol. 14, no. 7, pp. 2989–2998.

30. Pedregosa, F., Varoquaux, G., Gramfort, A., et al., Scikit-learn: machine learning in Python, *J. Mach. Learn. Res.*, 2011, vol. 12, pp. 2825–2830.

31. Ershov, E.B., Distribution of the determination coefficient to the general case of linear regression evaluated using various versions of the least squares method, *Ekon. Matem. Metody*, 2002, vol. 38, no. 3, pp. 107–120.

32. Theil, H., A rank-invariant method of linear and polynomial regression analysis, Parts I, II, III, *Proc. K. Ned. Akad. Wet., Ser. A: Math. Sci.*, 1950, vol. 53, pp. 386–392, 521–525, 1397–1412.

33. Sen, P.K., Estimates of the regression coefficient based on Kendall's tau, *J. Am. Stat. Assoc.*, 1968, vol. 63, pp. 1379–1389.

Translated by I. Obrezanova