

# Chapter 9

## Application of Machine Training Methods to Design of New Inorganic Compounds

**Nadezhda Kiselyova**

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

**Andrey Stolyarenko**

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

**Vladimir Ryazanov**

*A. A. Dorodnicyn Computing Centre of Russian Academy of Sciences, Russia*

**Oleg Sen'ko**

*A. A. Dorodnicyn Computing Centre of Russian Academy of Sciences, Russia*

**Alexandr Dokukin**

*A. A. Dorodnicyn Computing Centre of Russian Academy of Sciences, Russia*

### ABSTRACT

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

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

## INTRODUCTION

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

The present chapter is devoted to the use of precedent-based computer training methods for

searching for classification rules for inorganic substances and the application of these rules to predicting new compounds and evaluating their properties.

## STATEMENT OF THE PROBLEM OF DESIGNING NEW INORGANIC COMPOUNDS

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

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

22 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/application-machine-training-methods-design/69410](http://www.igi-global.com/chapter/application-machine-training-methods-design/69410)

## Related Content

---

### Visualizing Multi Dimensional Data

César García-Osorio and Colin Fyfe (2008). *Successes and New Directions in Data Mining* (pp. 236-276).

[www.irma-international.org/chapter/visualizing-multi-dimensional-data/29962](http://www.irma-international.org/chapter/visualizing-multi-dimensional-data/29962)

### Data Stream Mining Using Ensemble Classifier: A Collaborative Approach of Classifiers

Snehlata Sewakdas Dongre and Latesh G. Malik (2017). *Collaborative Filtering Using Data Mining and Analysis* (pp. 236-249).

[www.irma-international.org/chapter/data-stream-mining-using-ensemble-classifier/159506](http://www.irma-international.org/chapter/data-stream-mining-using-ensemble-classifier/159506)

### Estimating Semi-Parametric Missing Values with Iterative Imputation

Shichao Zhang (2010). *International Journal of Data Warehousing and Mining* (pp. 1-10).

[www.irma-international.org/article/estimating-semi-parametric-missing-values/44955](http://www.irma-international.org/article/estimating-semi-parametric-missing-values/44955)

### PSSRC: A Web Service Registration Cloud Based on Structured P2P and Semantics

Qian He, Baokang Zhao, Liang Chang, Jinshu Su and Ilun You (2016). *International Journal of Data Warehousing and Mining* (pp. 21-38).

[www.irma-international.org/article/pssrc/146851](http://www.irma-international.org/article/pssrc/146851)

### Sentiment Analysis of Game Review Using Machine Learning in a Hadoop Ecosystem

Arvind Panwar and Vishal Bhatnagar (2022). *Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines* (pp. 463-483).

[www.irma-international.org/chapter/sentiment-analysis-of-game-review-using-machine-learning-in-a-hadoop-ecosystem/308503](http://www.irma-international.org/chapter/sentiment-analysis-of-game-review-using-machine-learning-in-a-hadoop-ecosystem/308503)