Acta Crystallographica Section B
Structural Science
ISSN 0108-7681

Predicting regularities in lattice constants of $GdFeO_3$-type perovskites

Asifullah Khan[a]* and Syed Gibran Javed[b]

[a]Department of Information and Computer Sciences, Pakistan Institute of Engineering and Applied Sciences, Nilore, Islamabad, Pakistan, and [b] Faculty of Computer Sciences and Engineering, Ghulam Ishaq Khan (GIK) Institute of Engineering Science and Technology, Swabi, Pakistan

Correspondence e-mail: asif@pieas.edu.pk

Received 11 August 2007
Accepted 9 November 2007

A novel idea of employing genetic programming to obtain mathematical expressions representing the dependency of lattice constants (LC) on their atomic parameters is presented in this paper. The results obtained from simulations reveal that only two atomic parameters are sufficient for LC prediction of $GdFeO_3$-type perovskites. In addition, an advantage of this approach is that there is no need to save any trained model as in the case of other existing machine-learning based approaches.

1. Introduction

Traditional methods of determining lattice constants (LC) are usually based on X-ray, neutron or electron diffraction techniques. These techniques form a powerful approach and are very popular among crystallographers. On the other hand, theoretical approaches for determining LC are also gaining popularity because they do not require any experimental setup or sample materials. They use mathematical models and atomic features of the materials to determine LC, which are easy to simulate on computers.

LC prediction of crystalline materials is helpful in selecting an appropriate material in many industry-related problems, especially the material design of interface applicants (Liu & Edgar, 2002). Lattice mismatch between thin films and their substrates is a well known industrial issue (Dawber *et al.*, 2005) and it limits the large-scale production of thin films. Therefore, it is the requirement of industry to develop prediction models that can correctly predict the values of LC of unknown crystalline compounds with low temporal and computational cost. Lufaso & Woodward (2001), using a bond-valence method, have developed a software program named *SPuDS* for the prediction of crystal structures of perovskites. Previously, an approach based on support vector machine (SVM) has been used for efficiently predicting LC without losing the generalization performance (Javed *et al.*, 2007). Chonghe *et al.* (2003) have also proposed an approach based on artificial neural networks (ANN) for a similar kind of problem. However, because of their inherent inability, both approaches are unable to provide the functional form of the trained models as mathematical expressions. In this study, the capabilities of genetic programming (GP) are explored to automatically find mathematical expressions rather than memory-based models that can be used to predict the LC of novel $GdFeO_3$-type perovskites. For LC prediction of $GdFeO_3$-type perovskites, the developed expressions need only the ionic radii of the cations.

2. Experimental

*MATLAB*7 (Mathworks Inc., 2005) provides an easy-to-use computational environment with problems and solutions being expressed in familiar mathematical notations. The *GPLAB* (Silva, 2005) toolbox is used to employ the capabilities of GP in the proposed scheme. To check and verify the performance of the proposed methodology, we have used the data of orthorhombic $GdFeO_3$-type perovskite materials. The space group of the orthorhombic $GdFeO_3$ samples is *Pnma* (O'Keeffe & Hyde, 1977), while the coordination number for the A-site cation is 8. The dataset used in this work was taken from Chonghe *et al.* (2003) and the Inorganic Crystal Structure Database (ICSD,

2002). The set of ionic radii that we had used was from Lide (1999) and Chonghe *et al.* (2003). The whole dataset comprised a total of 161 samples, out of which 157 samples were used for training.. The remaining samples were used as a test dataset and were treated as new compounds. These novel data samples were taken from Chonghe *et al.* (2003), Hannerz *et al.* (1999), Woodward *et al.* (2000), Kim, Demazeau, Alonso *et al.* (2001) and Kim, Demazeau, Presniakov *et al.*, 2001.

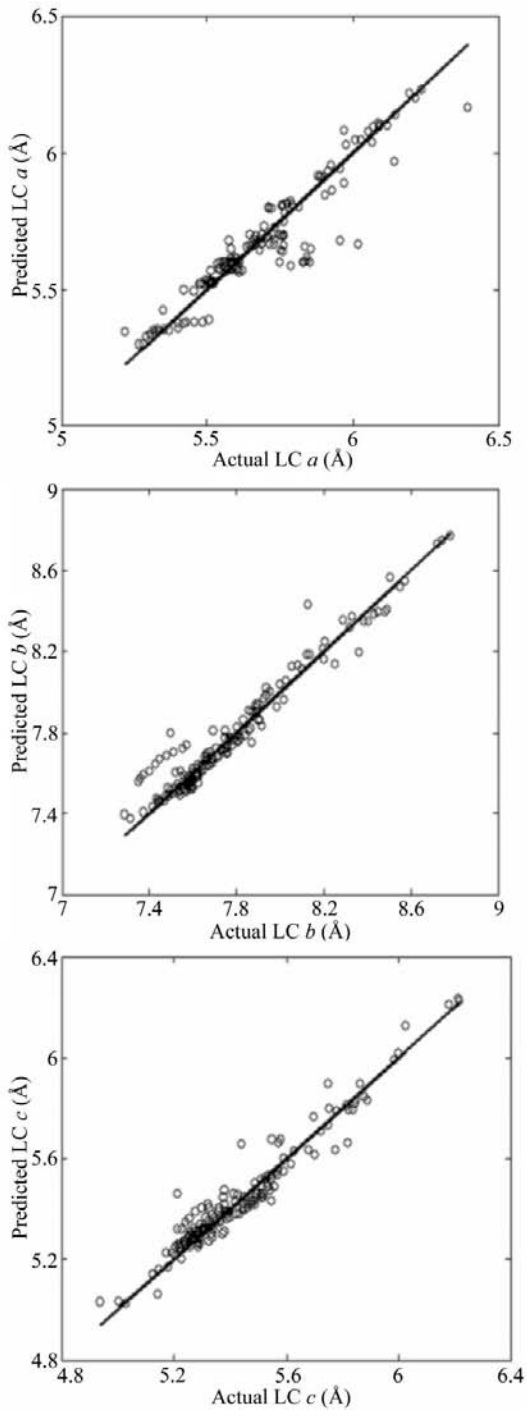

Figure 1
Actual *versus* predicted lattice constants *a*, *b* and *c* obtained with the GP approach. The solid trend lines represent perfect agreement.

3. Lattice constants and its dependencies

It is well known that the stability of an ionic arrangement depends upon Gibbs free energy. However, Gibbs free energy for crystals depends upon lattice energy, which is a function of the radii and valence of the constituent ions. In this study we are interested in predicting LC using only the ionic radii. We can express this relationship as

$$\mathrm{LC} = f(r_A, r_B, r_\mathrm{O}), \tag{1}$$

where r_A, r_B represent the ionic radii of cations A and B, respectively, and r_O = 1.32 Å is the ionic radius of the oxide ion. r_O remains the same for all the samples used in this study and thus can be neglected. Consequently

$$\mathrm{LC} = f(r_A, r_B). \tag{2}$$

The stability of perovskites being related to ionic radii is also expressed by the tolerance factor

$$t(r_A, r_B) = \frac{r_A + 1.32}{2^{1/2}(r_B + 1.32)}. \tag{3}$$

We assume that the inclusion of the tolerance factor in the prediction model for LC, as an independent variable, is helpful in quickly correlating LC with ionic radii. Consequently, the functional form of LC can be written as

$$\mathrm{LC} = f(r_A, r_B, t(r_A, r_B)). \tag{4}$$

4. Developing LC prediction models

Our proposed methodology of developing LC prediction models, representing the dependency of LC on its atomic parameters, mainly consists of three major modules. These consist of: dataset generation, GP expression evolution module and GP expression testing.

In dataset generation module, the whole dataset is partitioned into two subsets: training and testing sets. Each sample pattern in the training set consists of: input pattern, comprising the ionic radii of cation A, r_A, and B, r_B, and the tolerance factor t. The corresponding output comprises of experimental LC a, b or c. In GP expression evolution module, mathematical expressions are developed, which represent the functional form of the LC through an evolutionary process.[1]

Each candidate solution in the initial population represents the possible functional form of the LC prediction model. A solution evaluation operation takes the initial population and the training samples as an input and generates a set of output values, which are used to compute the corresponding fitness given by (5).

$$\mathrm{fitness} = \mathrm{PAD}/2 + (\mathrm{MSE} \times 100)/2, \tag{5}$$

where the percentage of absolute difference (PAD) and mean-square error (MSE) are given by the expressions

$$\mathrm{PAD}\ (\%) = \frac{|\mathrm{experimental} - \mathrm{predicted}|}{\mathrm{experimental}} \times 100 \tag{6}$$

and

$$\mathrm{MSE} = \frac{(\mathrm{experimental} - \mathrm{predicted})^2}{\mathrm{Total\,No.\ of\ samples}}. \tag{7}$$

[1] Supplementary data for this paper are available from the IUCr electronic archives (Reference: SO5010). Services for accessing these data are described at the back of the journal.

Table 1
Statistical summary of PAD on training and testing data.

	LC	Mean	Maximum	Minimum	Standard deviation
Training dataset	*a*	0.89	5.86	0.00	1.10
	b	0.64	3.97	0.00	0.74
	c	0.77	4.79	0.00	0.73
Testing dataset	*a*	1.31	2.77	0.21	1.19
	b	1.47	2.99	0.09	1.18
	c	0.84	1.15	0.34	0.37

Table 2
GP-predicted LC values and their corresponding PAD (%).

Compounds		$SrNbO_3$	$CaFeO_3$	$TlNiO_3$	$TlFeO_3$
Expt. LC	*a*	5.6944	5.352	5.3677	5.4465
	b	8.0864	7.539	7.562	7.7927
	c	5.6894	5.326	5.2549	5.3172
GP Pred. LC	*a*	5.7064	5.4482	5.3922	5.5974
	b	8.0941	7.637	7.4487	7.5597
	c	5.7521	5.3668	5.1946	5.2990
PAD %	*a*	0.21	1.79	0.46	2.77
	b	0.09	1.30	1.49	2.99
	c	1.10	0.76	1.14	0.34

The candidate with the best fitness value is checked against the termination criterion. The outcome of the GP module is thus a functional expression with the best fitness value. More details on the subject can be found in Banzhaf *et al.* (1998). Once the best GP expression is found, its performance on testing data is checked.

5. Results and discussion

The mathematical expressions (see supplementary data) for LC prediction obtained from the GP process were evaluated on the training dataset. Fig. 1 shows the plots of the experimental *versus* GP predicted values of LC *a*, *b* and *c*. It can be observed from these plots that the GP-predicted LC are concentrated near the ideal line and consequently, the GP model is able to learn the functional dependency of the LC on the atomic parameters. This fact can also be observed from Table 1, which shows the statistical summary of PAD (%) on the training dataset. The results in this table show that the average PAD is less than 0.9% for all the three lattice constants of the training data set.

Table 2 shows the GP-predicted LC and their corresponding PAD values for the four novel compounds used as testing dataset. The statistical summary of PAD (%) on the testing dataset is shown in Table 1.

It can be observed that the average PAD is less than 1.5% for all the three lattice constants of the testing dataset. According to Chonghe *et al.* (2003), the average PAD for these compounds using the ANN approach is less than 2%. On the other hand, the SVM-based method shows a more accurate prediction as the average PAD is less than 0.58% (Javed *et al.*, 2007). Based on these results, it can be concluded that with novel $GdFeO_3$ perovskites, GP performed better than ANN but not as well as SVM. However, it should be noted that GP uses only two atomic parameters, whereas both ANN- and SVM-based approaches have used five atomic parameters.

6. Conclusions

In this work GP is used to develop efficient prediction models for the LC of $GdFeO_3$-type perovskites. The experimental results obtained using the evolved expressions confirm the generality of the proposed method and show that the average PAD is approximately 1.5% in all cases. Comparing it to the other machine-learning-based techniques, the proposed technique is successful in presenting three functional expressions for LC *a*, *b* and *c* by using just two atomic parameters; the ionic radii of cations *A* and *B*. For accuracy enhancement in LC prediction, additional atomic parameters such as electronegativity, valence, bond lengths *etc.* may be exploited. The employment of the advanced machine-learning approach, GP, has considerable potential in optimization- and regression-oriented problems of crystallography.

The authors acknowledge the support of Dr Nasr A. Mohamed Hafz, Senior Research Scientist, Advanced Photonics Research Institute, GIST, South Korea, and Dr Abdul Majid, Associate Professor, Department of Computer and Information Sciences, PIEAS, Pakistan, during the course of this work.

References

Banzhaf, W., Nordin, P., Keller, R. E. & Francone, F. D. (1998). *Genetic Programming An Introduction: On the Automatic Evolution of Computer Programs and Its Applications.* San Francisco, California: Morgan Kaufmann Publishers, Inc.
Chonghe, L., Yihao, T., Yingzhi, Z., Chunmei, W. & Ping, W. (2003). *J. Phys. Chem. Solid*, **64**, 2147–2156.
Dawber, M., Rabe, K. M. & Scott, J. F. (2005). *Rev. Mod. Phys.* **77**, 1083–1131.
Hannerz, H., Svensson, G., Istomin, S. Y. & D'yachenko, O. G. (1999). *J. Solid State. Chem.* **147**, 421–428.
Inorganic Crystal Structure Database (2002). Fiz Karlsruhe, Germany, http://www.fiz-karlsruhe.de/icsd.html or http://icsdweb.fiz-karlsruhe.de.
Javed, S. G., Khan, A., Majid, A., Mirza, A. M. & Bashir, J. (2007). *Comput. Mater. Sci.* **39**, 627–634.
Kim, S. J., Demazeau, G., Alonso, J. A. & Choy, J. H. (2001). *J. Mater. Chem.* **11**, 487–492.
Kim, S. J., Demazeau, G., Presniakov, I. & Choy, J. H. (2001). *J. Solid State Chem.* **161**, 197–204.
Lide, D. R. (1999). *Handbook of Chemistry and Physics.* Boca Raton, Florida, London: CRC Press/Chapman and Hall.
Liu, L. & Edgar, J. H. (2002). *Mater. Sci. Eng. R*, **37**, 61–127.
Lufaso, M. W. & Woodward, P. M. (2001). *Acta Cryst.* B**57**, 725–738.
Mathworks Inc. (2005). *MATLAB*7.0. Mathworks Inc., Natick, Massachusetts, http://www.mathworks.com.
O'Keeffe, M. & Hyde, B. G. (1977). *Acta Cryst.* B**33**, 3802–3813.
Silva, S. (2005). *GPLAB*, http://gplab.sourceforge.net.
Woodward, P. M., Cox, D. E., Moshopoulou, E., Sleight, A. W. & Morimoto, S. (2000). *Phys. Rev. B*, **62**, 844.