06

Determination of the proportions of platinum atoms in agglomerates of bimetallic nanoparticles using machine learning methods

© Ya.N. Gladchenko-Djevelekis, D.B. Tolchina, S.V. Belenov, V.V. Srabionyan, V.A. Durymanov, I.A. Viklenko, L.A. Avakyan, A.A. Alekseenko, L.A. Bugaev

Southern Federal University,
344090 Rostov-on-Don, Russia
e-mail: ygl@sfedu.ru; vvsrab@sfedu.ru

Received October 24, 2024
Revised March 11, 2025
Accepted April 1, 2025

In this paper, we consider the applicability of machine learning methods, in particular, artificial neural networks, to obtain information on the distribution of target substance atoms in aggregates of bimetallic nanoparticles of various architectures. To solve the problem, we use data on paired radial distribution functions of atoms, the direct sources of which are experimental methods of X-ray diffraction and X-ray absorption spectroscopy from an extended energy region of the spectrum. The trained model of the artificial neural network demonstrates high accuracy in determining the proportions of platinum atoms in the composition of nanoparticles of various architectures in the agglomerate (determination coefficient ~ 0.98). To verify the trained model, experimental data for catalysts containing bimetallic PtCu nanoparticles were used. Verification showed a high generalisability of the model, which indicates the promising application of this approach to the determination of platinum consumption efficiency in the synthesis of platinum-containing nanoparticle-based catalysts.

Keywords: Core shell nanoparticles, gradient nanoparticles, RDF, EXAFS, artificial neural networks, CatBoost.

DOI: 10.61011/TP.2025.08.61736.365-24

Introduction

Presently, platinum catalyzers are applied in many industries. In particular, platinum-containing nanoparticles (NP) applied to a finely-dispersed carbon material are actively used when designing fuel cells as catalyzers [1]. Due to high price of platinum, great attention is paid to designing high-effective and stable platinum-containing catalyzers with lower cost [2–6]. One of the ways of solving this problem includes using bimetal platinum-containing nanoparticles with addition of *d*-metals, for example, nickel, cobalt or copper [1,7]. At the same time, catalytic properties of these nanoparticles depend both on a composition and their architecture, i.e. spatial distribution of the components across the nanoparticle volume.

It is known that the bimetal nanoparticles with a core-shell architecture with the platinum shell demonstrate catalytic activity that can be compared with single-metal platinum nanoparticles, but contain much less expensive platinum [2–4]. However, it should be noted that during operation the architecture of these nanoparticles can change in time. For example, a boundary of transition from the core to the shell can be less pronounced due to thermal effects [8,9], which in case of a thin layer of the shell can result in deterioration of the catalytic properties of the nanoparticles. The recent studies have shown that a smooth transitive layer between the core and shell areas made it possible to achieve improved stability of a catalyst [10,11]. It was proposed to name this architecture of the nanoparticles a „gradient“ one [10,12,13].

Despite the entire promising potential of the „gradient“ architecture of the nanoparticles, a process of multi-stage synthesis can include formation of the nanoparticles of other types, whether it is single- or bi-metal nanoparticles of various architectures. That is why production of the nanoparticles of a pre-defined architecture with certain properties necessitates significant optimization of the synthesis conditions, which requires availability of structural information on the nanoparticles. The most common sources for obtaining this information include X-ray absorption spectroscopy (XAS), methods of X-ray diffraction (XRD) and high-resolution electron microscopy with elementary mapping and some other methods [14,15].

Application of machine learning methods can significantly simplify and accelerate extraction and analysis of the structural information on nanoobjects. For example, it was demonstrated that using the neuron networks can extract information on coordination numbers and distances from XAS data [16,17] and with high accuracy it can determine the architecture of the nanoparticles from data on paired radial distribution functions of atoms (PRDF) [18].

The present study is dedicated to investigating a fundamental determinability of efficiency of platinum consumption during synthesis, i.e. determination of a proportion of platinum, which is consumed for forming the nanoparticle with a specific architecture.

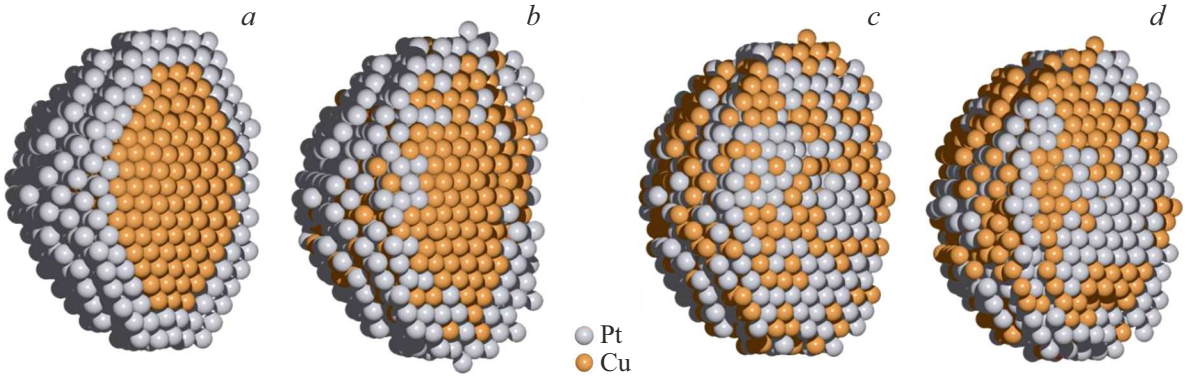

Figure 1. Considered architectures of the bimetal nanoparticles: *a* — core-shell, *b* — „gradient“, *c* — disordered and *d* — aggregated alloys.

1. Methods and approaches

1.1. Sources for obtaining PRDF

The main methods for obtaining PRDF are analysis of data of X-ray diffraction and X-ray absorption spectroscopy from the extended energy region of the spectrum (EXAFS). The method of extraction of PRDF from XRD is based on a dependence of the radial distribution function of the atomic density on intensity of coherent scattering of X rays during diffraction [19]. Due to physical limitations of measurement of scattering of X rays, extraction of the radial distribution functions from XRD originates edge effects that are manifested in appearance of false maximums in curves of the distribution function. Extraction of data on PRDF from the EXAFS data is based on multiparameter optimization of the Fourier transform F(R) of an initial signal. For the specimen-average representative nanoparticle, this analysis allowed determining radii of coordination spheres (~ 0.01 Å) [20] with high accuracy and coordination numbers (the error $\sim 10\,\%$) with much less accuracy as well as a parameter of temperature and structure disordering [21,22]. It should be separately noted that a EXAFS-analysis process can substantially differ depending on a studied material and that reliability of determining structure parameters decreases when considering more remote coordination spheres. The present study uses PRDFs directly calculated for molecular nanoclusters. Therefore, when comparing with experimental data, correctness of the results necessitates additional normalization of the coordination numbers determined from the EXAFS analysis. It is due to the fact that metal atoms can be in two states belonging to two components: nanoparticles and an oxide:

$$N_{A-B}^{Np} = \frac{V}{V - N_{A-O}^{\mathrm{EXAFS}}} \cdot N_{A-B}^{\mathrm{EXAFS}}, \qquad (1)$$

where N_{A-B}^{EXAFS} — the coordination number of the atoms of the sort B in relation to the sort A, as obtained from the EXAFS-analysis; V — the assumed coordination number for atoms of the metal of the sort A in an oxidized state, which can be in the material not in the composition of the nanoparticle (it is assumed in the present paper that $V = 6$); N_{A-O}^{EXAFS} — the coordination number of the oxygen atoms in relation to the atoms of the sort A. The expression (1) is derived with more details in Appendix 1.

1.2. Machine learning methods. Introduction to the problem, used metrics

The set problem for determining efficiency of platinum consumption during synthesis of the nanoparticles of the pre-defined architecture can be reduced to determining the proportions of platinum that were consumed for formation of the nanoparticles of the various architectures in their agglomerate. Probability of formation of a specific architecture of the nanoparticle substantially depends on a synthesis procedure [6,9–11,23] and, therefore, we limited ourselves to consideration of the most probable nanoparticle architectures formed during synthesis of the platinum-containing catalysts. Thus, the present study has considered single-metal nanoparticles of platinum and bimetal nanoparticles of the PtCu composition with the following architectures: the copper core and the platinum shell, the platinum-copper with the structure of the alloy and the aggregated alloy and the gradient nanoparticle with the copper core and the platinum shell. Images of the considered architectures of the bimetal nanoparticles are shown in Fig. 1.

Thus, the set problem is reduced to a problem of multipurpose regression on tabular data, which can be solved by the machine learning methods with a teacher. Formally, the model operation can be presented as follows:

$$\Phi(x^i, \Theta) \to \left\{ c_{\mathrm{Pt}}, c_{\mathrm{PtCu}}, c_{\mathrm{Cu@Pt}}, c_{\mathrm{PtCu_{aggr}}}, c_{\mathrm{CuPt_{grad}}} \right\}^i,$$

where x^i — the PRDF for i-th data copy, Θ — the array of the model parameters, $\{c_{\mathrm{Pt}}, c_{\mathrm{PtCu}}, c_{\mathrm{Cu@Pt}}, c_{\mathrm{PtCu_{aggr}}}, c_{\mathrm{CuPt_{grad}}}\}$ — the model-predicted vector of the platinum proportions in the composition of the nanoparticles of the considered architectures in the agglomerate. Teaching of the model includes selection

of values of the parameter array Θ in such a way as to approximate the model-predicted and the reference values of the proportion of the platinum atoms as per a certain loss function. In the present study, the loss function was a function of the mean squared error (MSE) that is determined by the expression:

$$MSE = \frac{1}{n}\Sigma_{i=1}^{n}(y_i - y_i^*)^2,$$

where y_i and y_i^* are the real and the model-predicted vector of target values, respectively.

2. Results and discussion

2.1. Formation of a learning data set

The present study considers PRDFs of the platinum single-metal and platinum-copper bimetal nanoparticles of a spherical form with the four different architectures (Fig. 1). The atomic models of the nanoparticles were obtained by cutting the spherical area from an infinite FCC crystal with subsequent molecular-dynamic simulation, as described in the paper [18]. Totally, we obtained 1456 different platinum-copper bimetal nanoparticles with sixteen different sizes (from 1.3 to 6 nm) and thirteen ratios of the Pt:Cu components (from 20 : 80 to 80 : 20) and 22 platinum single-metal nanoparticles of the sizes from 1.4 to 9.5 nm.

The assembly of the five nanoparticles of the different architectures is characterized by PRDF obtained by the formula (2), where c_{arch} — the proportion of a contribution of the nanoparticles of a given architecture in the assembly, n_{arch}^{Pt} — the number of the platinum atoms in the nanoparticle of a given architecture, PRDF_{arch} — PRDF of the specific nanoparticle of the pre-defined architecture

$$\mathrm{PRDF} = \frac{\Sigma_{arch} c_{arch} \cdot \mathrm{PRDF}_{arch} \cdot n_{arch}^{\mathrm{Pt}}}{\Sigma_{arch} c_{arch} \cdot n_{arch}^{\mathrm{Pt}}}. \qquad (2)$$

To form a data set of this type, it is necessary to randomly generate a set of concentrations carch, whose sum will be unity within one assembly. These coefficient were generated by means of the Dirichlet function with the parameters $\alpha_1 = [0.6, 0.6, 0.6, 0.6, 0.6]$. The distribution of carch obtained as a result of this operation is shown in Fig. 2, *a*. The obtained data have a high density in the neighborhood of the point A , which is a proportion of platinum in the composition of the single-metal platinum nanoparticle (Fig. 2, *b*). It is related to the fact that we have a limited set of single-metal nanoparticles with a high average value of the number of the platinum atoms in the nanoparticles. When summing the PRDFs (the formula (2)) of the nanoparticles of the different architectures, with greater probability we select large single-metal platinum nanoparticles. Accordingly, the proportion of platinum, which is „consumed" for their formation is larger than all others. This distribution was corrected by combining two data sets that obtained as a result of generation of the coefficients c_{arch} by means of the Dirichlet functions with the parameters $\alpha_2 = [0.2, 0.6, 0.6, 0.6, 0.6]$ and $\alpha_3 = [0.5, 3, 3, 3, 3]$ (Fig. 2, *c*).

For each set of the five concentrations carch, 208 nanoparticles of each architecture with the different sizes and the different ratios of the Pt : Cu components are randomly selected and the assembly PRDF is calculated by the formula (2). Thus, a learning sample is formed and totally consists of 208 000 lines.

The data, which are of interest to us, on the proportions of the platinum atoms, which are consumed to form the nanoparticle of a certain architecture in the assembly, c_{arch}^{Pt}, is determined as

$$c_{arch}^{\mathrm{Pt}} = \frac{c_{arch} \cdot n_{arch}^{\mathrm{Pt}}}{\Sigma_{arch} c_{arch} \cdot n_{arch}^{\mathrm{Pt}}}. \qquad (3)$$

The resulted distribution of c_{arch}^{Pt} is shown in Fig. 2, *d*. The resultant set of the learning data was divided into 3 parts: the coaching set — 60 % of the entire sample; the validation set — 15 % of the sample and the test set — 25 % of the sample. According to the common paradigm, the coaching data set was used for training the parameter matrix Θ of the used machine learning models, the validation sample was used to select hyperparameters of the models and to determine a time when training stops and the test sample was used for unbiased evaluation of the quality of model operation.

2.2. Selection of models

When constructing the matrix of correlations of the input data, it was found that there were significant correlations between many features. These correlations are caused by a structure of the input structure, which is Gaussian peaks. There is multicollinearity of the data and it imposes limitations on selection of possible models or pre-processing of the data. Fig. 3 shows the matrix of correlations between the input and output parameters.

Taking into account specific features of the input data, a ridge regression model was used as the basic machine learning model. High average values of the determination coefficient were achieved for it, and they were ~ 0.98 on average. However, when checking the model on the experimental data, the predicted proportions of the platinum atoms in agglomerates of the bimetal nanoparticles have no physical meaning. Besides, the distribution of residuals contained significant ejections. In this regard, a perceptron model was considered and it realized linear regression with additional application of a sigmoid function to the output data. This enabled to avoid negative values when predicting the target parameters. But at the same time the values of the determination coefficient dropped to 0.58, 0.23 and 0.12 for the models that determine the PROP ORTION of the platinum atoms in the composition of the disordered, aggregated and „gradient" nanoparticles, respectively. Fig. 4 shows the matrix of the perceptron model coefficients with

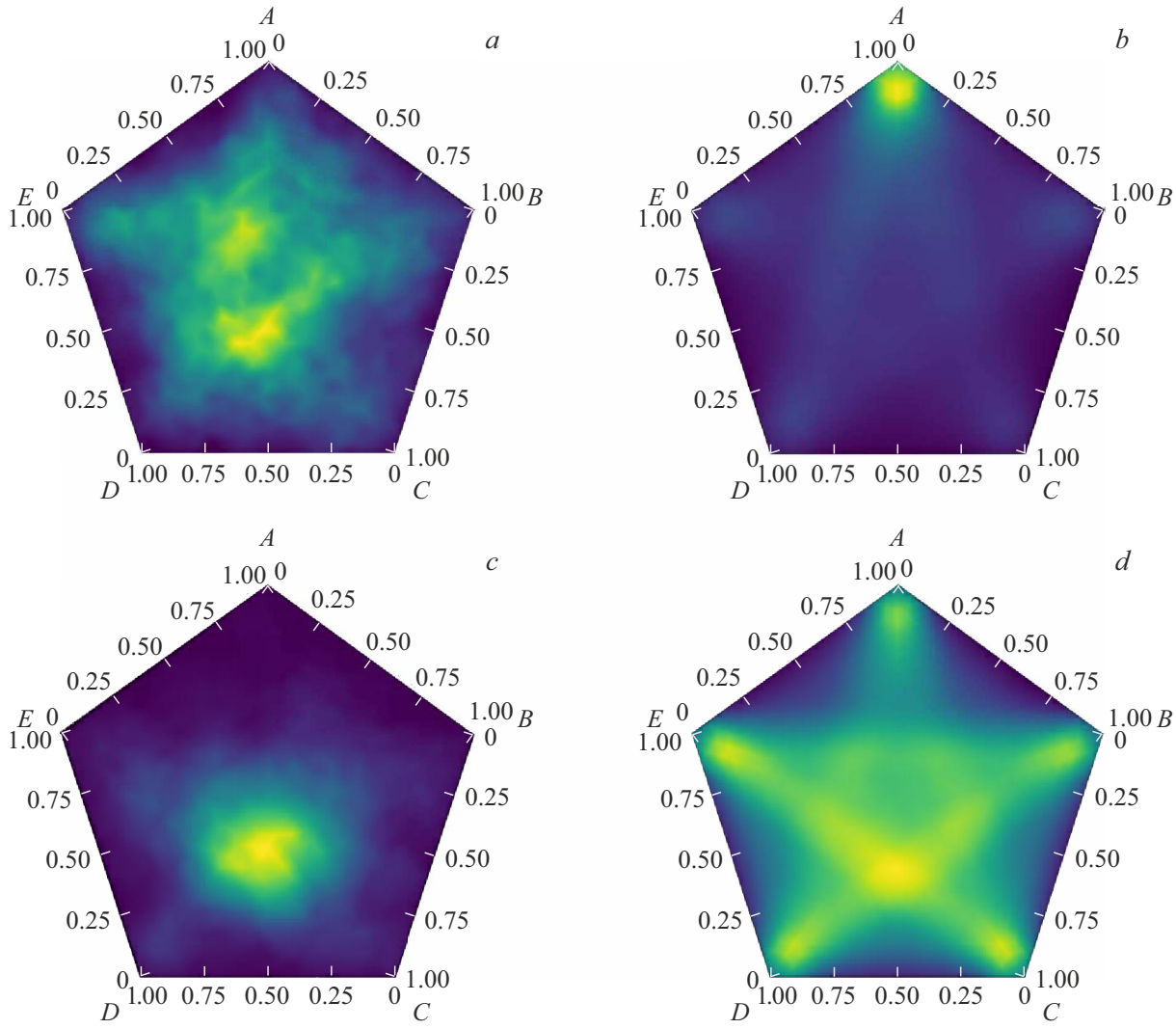

Figure 2. a, c — the distribution of proportions of the nanoparticles of the various architectures; b, d — the distribution of proportions of the platinum atoms in an assembly the five nanoparticles. A — Pt, B — PtCu, C — Cu@Pt, D — $PtCu_{aggr}$, E — $CuPt_{grad}$.

a form of the typical paired radial distribution functions of atoms for these architectures of the nanoparticles.

It is known than the models based on gradient boosting on solving trees can demonstrate both high accuracy and high generalizability when operating the tabular data [24]. Therefore, in this study we limited ourselves to consideration of the gradient boosting and artificial neuron networks (ANN) when solving the set problem.

The model of gradient boosting on the solving trees was CatBoostRegressor (CBR) realized in the CatBoost library [25,26], while an alternative approach was a fully connected neuron network with the architecture of Fig. 5. Details of tuning the hyperparameters of the used models are given in Appendix 2.

The described model of the neuron network receives a vector of the four serial PRDFs at the input: Pt-Pt, Pt-Cu, Cu-Pt, Cu-Cu and after that the data subsequently pass through the fully connected layers with an activation function ReLU and a regularization layer Dropout, which zeros 25 % of the output values. At the last layer, after the activation function ReLU, the function Softmax is applied to renormalize the output vector so as a sum of its components is equal to unity.

The model based on gradient boosting consisted of the solution trees of the depth of 10 and a number of boosting steps, which was 1000, the L2 regularization coefficient was 3 and the loss function was MultiRMSE [26].

It is known that some machine learning methods can be sensitive to a scale of the input data or presence of correlation. Therefore, when coaching the model, additional standardization of the input data was considered to bring them to a unified scale, so was application of a principal component analysis method to reduce the dimensionality of the input data and to obtain uncorrelated input features. However, these modifications of the input data have not resulted in any noticeable improvement of the quality of operation of ANN or the gradient boosting model in relation to the set problem.

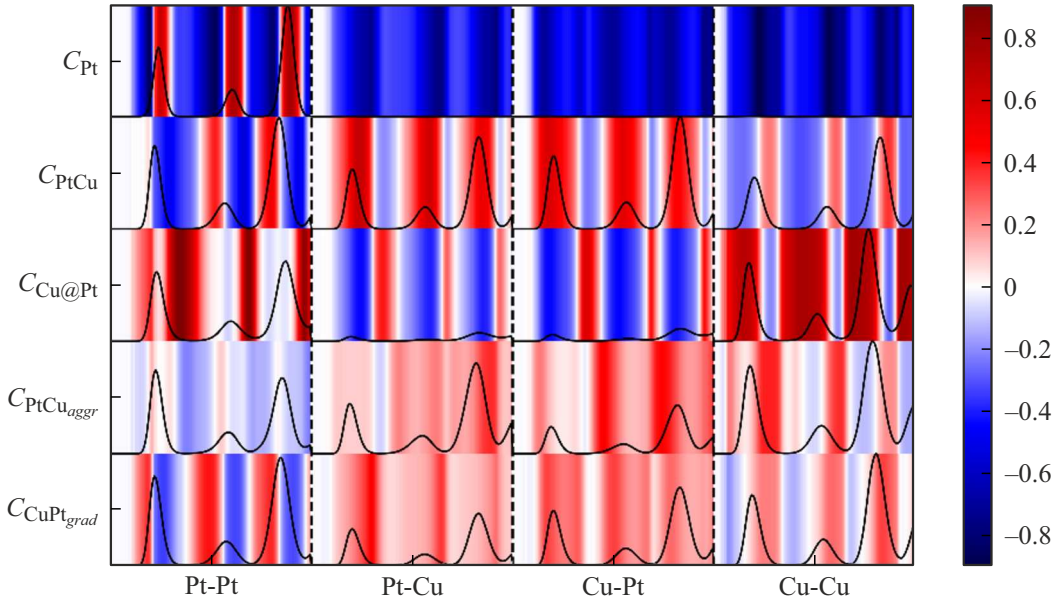

Figure 3. Matrix of correlations between the input and output parameters. Each architecture has a typical set of paired radial distribution functions of atoms, which is the input data.

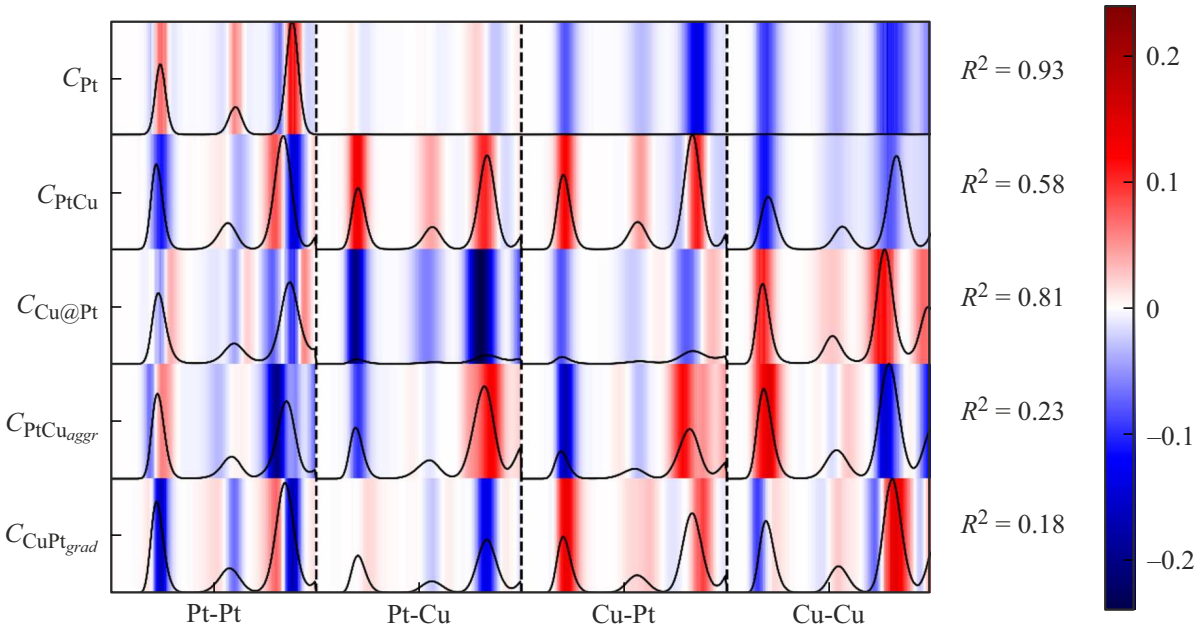

Figure 4. Matrix of the perceptron model coefficients. Each architecture has a typical set of paired radial distribution functions of atoms, which is the input data.

3. Learning results

3.1. Synthetic data

The results of operation of the trained models on the synthetic test data are shown in Fig. 6, while the error distribution statistics is shown in Fig. 7. Comparison of distribution of residuals for the ANN models that are trained on the data set corresponding to the generation parameters α_1 and the extended data set are shown in Fig. 8.

Even though both the models demonstrate an almost identical quality of description of the synthetic data, when comparing with the experimental data, the ANN shows systematically higher generalizability.

3.2. Comparison with the experimental data

In order to check applicability of the trained models based on ANN and CBR to the real data, consideration was additionally given to results of the study [10], in

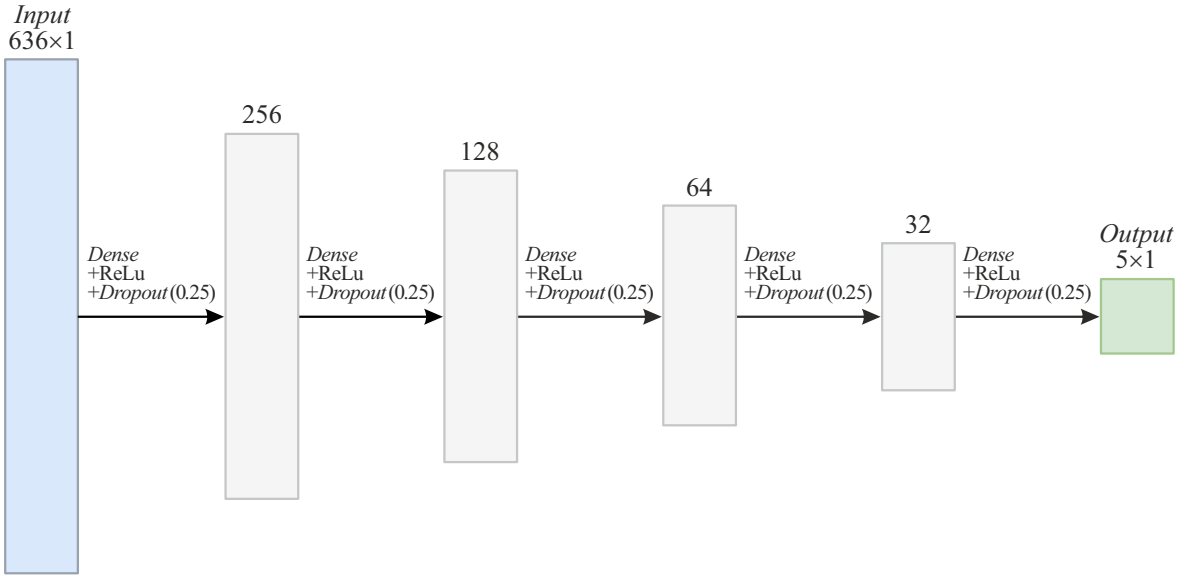

Figure 5. Diagram of the neuron network.

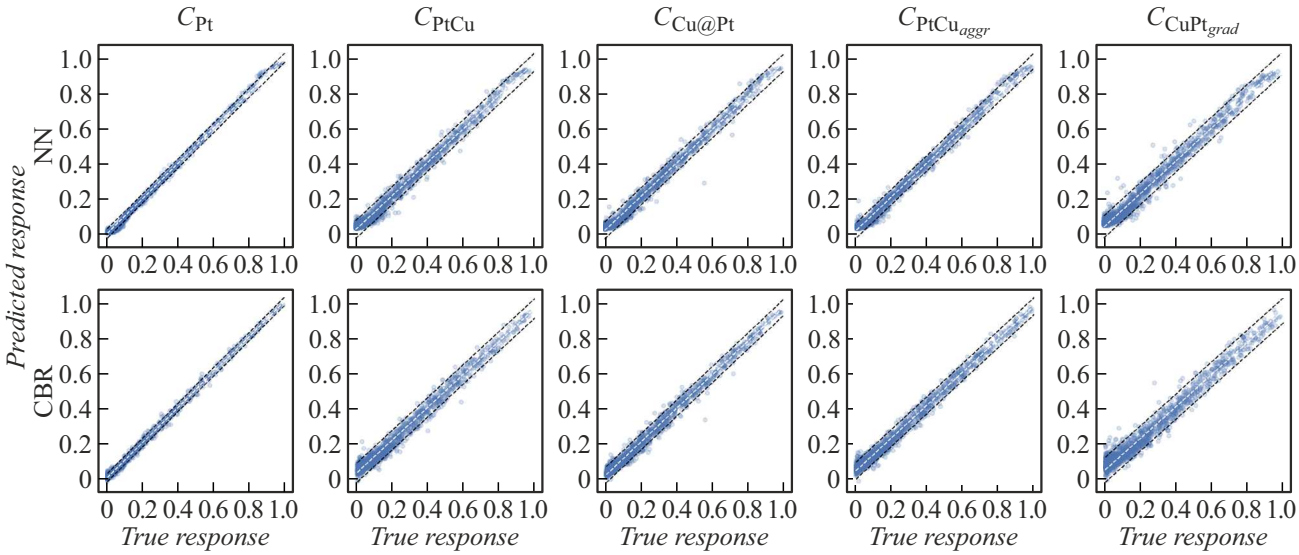

Figure 6. Comparison of the true and predicted values of the proportions of the platinum atoms in the assembly of the five nanoparticles for ANN (NN) and CBR.

which the authors used a multi-stage synthesis procedure for producing the „gradient“ platinum-copper nanoparticles. The synthesis procedure is schematically presented in Fig. 9 borrowed from the article [10].

The materials produced at the 2-nd, 3-d and 4-th stages are denoted as „PtCu_stage2“, „PtCu_stage3“ and „PtCu_stage4“ respectively. Besides, the study by S.V. Belenov et al. [11] has additionally investigated both simultaneous deposition of the atoms of platinum and copper with expected formation of the nanoparticles with the solid solution architecture, which are designated as „PtCu_sim“ and subsequent two-stage deposition of copper and platinum with expected formation of the particles with the copper core and the platinum shell, which were designated as „PtCu_seq“.

For correct operation of the trained models, these PRDFs obtained from the EXAFS analysis in the papers [10,11] were normalized according to the formula (1). The results of operation of the models are shown in Fig. 10.

It can be seen from the given results that the CBR model indicates presence of the significant proportion of the Pt atoms in the composition of the single-metal nanoparticles in all the studied specimens. On the other hand, the ANN model indicates an insignificant proportion of the single-metal Pt nanoparticles in the PtCu_sim and PtCu_seq specimens. For the PtCu_sim specimen, the CBR indicates

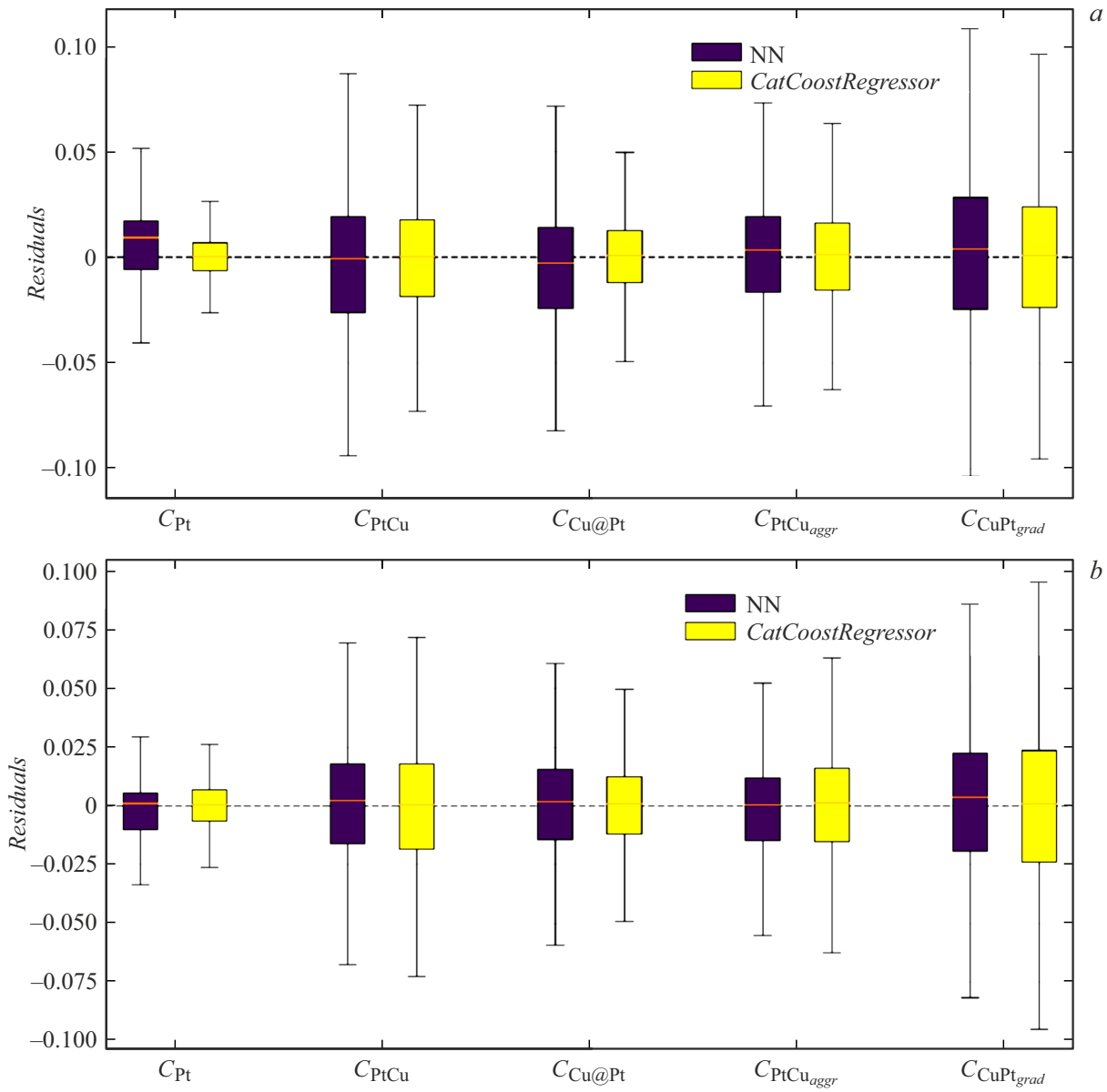

Figure 7. Comparison of distribution of residuals for ANN (NN) and CBR: *a* — the models trained on the data set that corresponds to the generation parameters α_1; *b* — the models trained on the extended data set.

an approximately uniform content of the Pt atoms in the composition of the considered architectures, except for the architecture of the disordered alloy, for which the model predicts the lesser Pt content. The architectures of the disordered alloy may really turn out to be unstable, since the FCC structures of the Pt and Cu volume specimens are characterized by clearly different lattice parameters, thereby making the aggregated-alloy architecture more probable in relation to the architecture of the disordered solution. Therefore, it is expected that with subsequent deposition of the components in the PtCu_seq specimen probability of formation of the nanoparticles with the disordered-alloy architecture will be not higher as compared to simultaneous deposition of the components in the PtCu_sim specimen. However, according to the results of using the CBR model, the proportion of the Pt atoms in the composition of the nanoparticles of the disordered-alloy architecture (32 %) is the greatest and significantly exceeds the similar value for PtCu_sim. At the same time, for the PtCu_seq specimen the CBR model indicates relative smallness of the proportion of the Pt atoms in the composition of the nanoparticles with the architectures „core-shell“ and „gradient“. And taking into account the synthesis procedure and the above-discussed results, it indicates insufficient generalizability of the CBR model in relation to the experimental data.

At the same time, for the PtCu_sim specimen the ANN model indicates that the large portion of the platinum atoms ($\sim$ 88 %) is consumed for forming the nanoparticles with

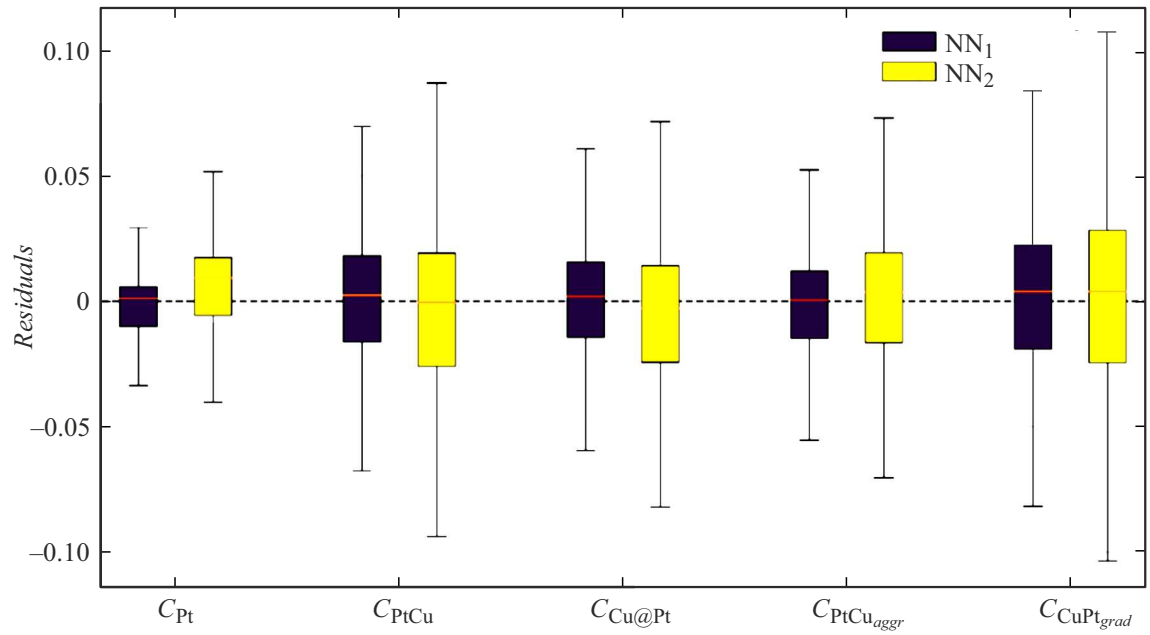

Figure 8. Comparison of distribution of residuals for NN_1 (the model trained on the extended data set) and NN_2 (the model trained on the data set that corresponds to the generation parameters α_1).

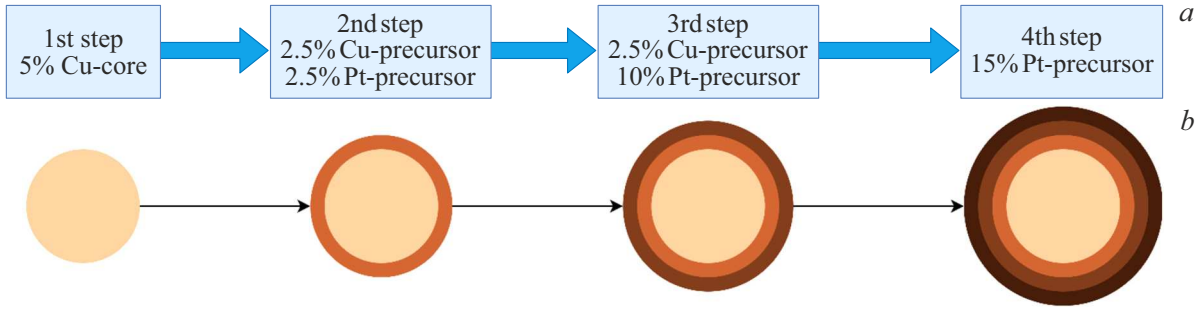

Figure 9. *a* — a diagram of step-by-step synthesis of the PtCu/C materials; *b* — a schematic image of the architecture of the PtCu nanoparticles that are formed in each of the four stages of successive synthesis of the PtCu/C gradient catalyst.

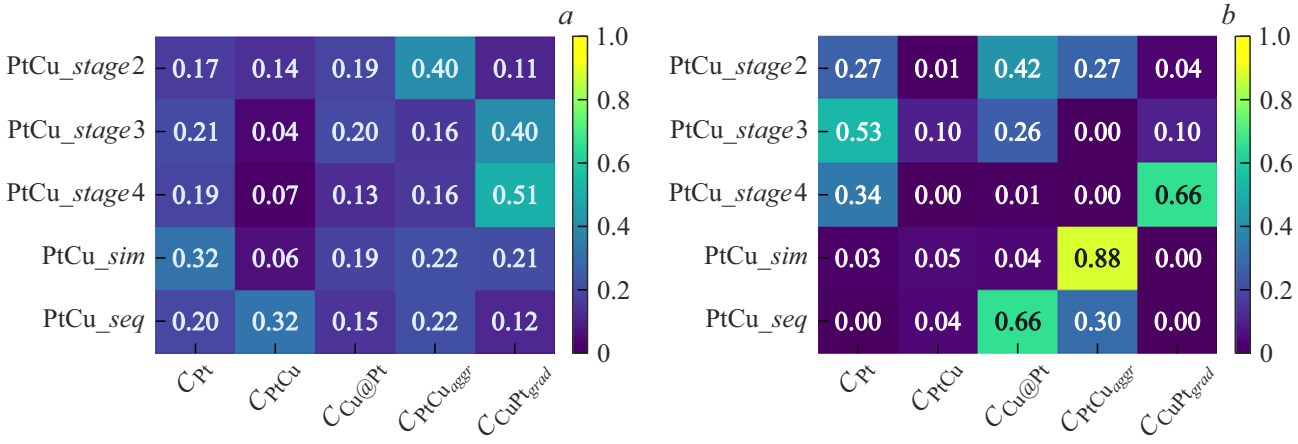

Figure 10. Result of use of the CBR model (*a*) and the ANN model (*b*) on the experimental data.

the aggregated-alloy architecture. In the PtCu_seq specimen, the large portion of platinum is in the composition of the nanoparticles with the core-shell architecture ($\sim 66\,\%$) and the significant portion thereof ($\sim 30\,\%$) is in the composition of the nanoparticles with the aggregated-alloy architecture. For the specimens produced at the 2-nd, 3-d and 4-th synthesis stages, the ANN model indicates monotonic increase of the proportion of the Pt atoms in the composition of the „gradient" architecture with simultaneous monotonic decrease of the nanoparticles with the „core-shell" architecture. At this, these synthesis stages exhibit a significant contribution of the single-metal platinum nanoparticles. The observed results of application of the ANN model for all the considered specimens are expected and logical and fully comply with the results of the study [10,11].

Conclusion

The study has demonstrated the fundamental determinability of the proportions of the atoms of a target substance in the composition of the aggregates of the nanoparticles of the various architectures using the machine learning methods, in particular, the artificial neuron networks, as per data of the paired radial distribution functions of atoms. The trained ANN model demonstrates high accuracy of determination of the proportions of the platinum atoms in the composition of the nanoparticles of the various architectures in the agglomerate with the determination coefficient R2 of more than 0.98 and the standard error deviation of 2.6 %. The trained model has been tested on the experimental data to show its high generalizability, thereby indicating the prospects of application of this approach to determination of efficiency of platinum consumption when synthesizing the platinum-containing nanoparticle-based catalysts.

Funding

The study was financially supported by the Russian Science Foundation (the project 23-21-00526 https://rscf.ru/project/23-21-00526/) in the Southern Federal University.

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1] K. Kodama, T. Nagai, A. Kuwaki, R. Jinnouchi, Y. Morimoto. Nat. Nanotechnol., **16** (2), 140 (2021). DOI: 10.1038/s41565-020-00824-w

[2] A.A. Alekseenko, A.S. Pavlets, S.V. Belenov, O.I. Safronenko, I.V. Pankov, V.E. Guterman. Appl. Surf. Sci., **595**, 153533 (2022). DOI: 10.1016/j.apsusc.2022.153533

[3] S. Hussain, H. Erikson, N. Kongi, A. Sarapuu, J. Solla-Gullón, G. Maia, A.M. Kannan, N. Alonso-Vante, K. Tammeveski. Int. J. Hydrogen Energy, **45** (56), 31775 (2020). DOI: 10.1016/j.ijhydene.2020.08.215

[4] A. Hrnjic, A.R. Kamšek, A. Pavlišič, M. Šala, M. Bele, L. Moriau, M. Gatalo, F. Ruiz-Zepeda, P. Jovanovič, N. Hodnik. Electrochim. Acta, **388**, 138513 (2021). DOI: 10.1016/j.electacta.2021.138513

[5] S. Zaman, L. Huang, A.I. Douka, H. Yang, B. You, B.Y. Xia. Angew. Chemie Int. Ed., **60** (33), 17832 (2021). DOI: 10.1002/anie.202016977

[6] M. Shao, Q. Chang, J.-P. Dodelet, R. Chenitz. Chem. Rev., **116** (6), 3594 (2016). DOI: 10.1021/acs.chemrev.5b00462

[7] W. Yan, D. Zhang, Q. Zhang, Y. Sun, S. Zhang, F. Du, X. Jin. J. Energy Chem., **64**, 583 (2022). DOI: 10.1016/j.jechem.2021.05.003

[8] M. Heinz, V.V. Srabionyan, L.A. Avakyan, A.L. Bugaev, A.V. Skidanenko, S.Y. Kaptelinin, J. Ihlemann, J. Meinertz, C. Patzig, M. Dubiel, L.A. Bugaev. J. Alloys Compd., **767**, 1253 (2018). DOI: 10.1016/j.jallcom.2018.07.183

[9] S. Belenov, A. Alekseenko, A. Pavlets, A. Nevelskaya, M. Danilenko. Catalysts, **12** (6), 638 (2022). DOI: 10.3390/catal12060638

[10] A.A. Alekseenko, V.E. Guterman, S.V. Belenov, V.S. Menshikov, N.Y. Tabachkova, O.I. Safronenko, E.A. Moguchikh. Int. J. Hydrogen Energy, **43** (7), 3676 (2018). DOI: 10.1016/j.ijhydene.2017.12.143

[11] S.V. Belenov, V.E. Guterman, N.Y. Tabachkova, E.A. Moguchikh, A.A. Alekseenko, V.A. Volochaev, N.M. Novikovskiy. Russ. J. Electrochem., **54** (12), 1209 (2018). DOI: 10.1134/S1023193518130062

[12] K. Boldt, S. Bartlett, N. Kirkwood, B. Johannessen. Nano Lett., **20** (2), 1009 (2020). DOI: 10.1021/acs.nanolett.9b04143

[13] X. Lyu, Y. Jia, X. Mao, D. Li, G. Li, L. Zhuang, X. Wang, D. Yang, Q. Wang, A. Du, X. Yao. Adv. Mater., **32** (32), 2003493 (2020). DOI: 10.1002/adma.202003493

[14] S. Mourdikoudis, R.M. Pallares, N.T.K. Thanh. Nanoscale, **10** (27), 12871 (2018). DOI: 10.1039/C8NR02278J

[15] L.J. Moriau, A. Hrnjić, A. Pavlišiš, A.R. Kamšek, U. Petek, F. Ruiz-Zepeda, M. Šala, L. Pavko, V.S. Šelih, M. Bele, P. Jovanovič, M. Gatalo, N. Hodnik. IScience, **24** (2), 102102 (2021). DOI: 10.1016/j.isci.2021.102102

[16] J. Timoshenko, D. Lu, Y. Lin, A.I. Frenkel. J. Phys. Chem. Lett., **8** (20), 5091 (2017). DOI: 10.1021/acs.jpclett.7b02364

[17] J. Timoshenko, A.I. Frenkel. ACS Catal., **9** (11), 10192 (2019). DOI: 10.1021/acscatal.9b03599

[18] L. Avakyan, D. Tolchina, V. Barkovski, S. Belenov, A. Alekseenko, A. Shaginyan, V. Srabionyan, V. Guterman, L. Bugaev. Comput. Mater. Sci., **208**, 111326 (2022). DOI: 10.1016/j.commatsci.2022.111326

[19] E. Collet, M. Buron, H. Cailleau, M. Lorenc, M. Servol, P. Rabiller, B. Toudic. X-ray diffraction for material science, in: UVX 2008–9e Colloq. Sur Les Sources Cohérentes Incohérentes UV, VUV X Appl. Développements Récents, EDP Sciences, Les Ulis, France, 2009, p. 21. DOI: 10.1051/uvx/2009005

[20] L.A. Bugaev, L.A. Avakyan, V. V. Srabionyan, A.L. Bugaev. Phys. Rev. B, **82** (6), 064204 (2010). DOI: 10.1103/PhysRevB.82.064204

[21] D.C. Koningsberger, B.L. Mojet, G.E. van Dorssen, D.E. Ramaker. Top. Catal., **10** (3/4), 143 (2000). DOI: 10.1023/A:1019105310221
[22] J.A. van Bokhoven, C. Lamberti. X-Ray Absorption and X-Ray Emission Spectroscopy (John Wiley & Sons, Ltd, Chichester, UK, 2016), DOI: 10.1002/9781118844243
[23] R. Wang, H. Wang, F. Luo, S. Liao. Electrochem. Energy Rev., **1** (3), 324 (2018). DOI: 10.1007/s41918-018-0013-0
[24] V.V. Kitov. Stat. Econ., **4**, 22 (2016). DOI: 10.21686/2500-3925-2016-4-22-26
[25] L. Prokhorenkova, G. Gusev, A. Vorobev, A.V. Dorogush, A. Gulin. (2017). http://arxiv.org/abs/1706.09516
[26] Electronic source. Available at: https://catboost.ai/en/docs/

Translated by M.Shevelev

Appendix 1

Derivation of the formula (1)

The equation associates the metal's coordination numbers (CN) obtained from EXAFS with the coordination numbers in the material with taking into account the fact that the metal atoms can be in the two states belonging to the two components: the nanoparticles (hereinafter referred to as NP) and the oxide (hereinafter referred to as Ox).

The coordination numbers are designated as N, the phase to which it belongs is designated by a superscript in brackets $N^{(\mathrm{NP})}$, $N^{(\mathrm{Ox})}$. The subscript contains information on the pair of atoms M $-$X, the central atom M — the surrounding atom X (M = Cu, Pt, X = Cu, Pt, O). The coordination number is not an additive magnitude along the states of atoms, but it can be expressed via the additive ones — the number of bonds $n_{\mathrm{M\text{-}X}}$ and the number of atoms n_{M} as

$$N_{\mathrm{M}-X} = \frac{n_{\mathrm{M\text{-}X}}}{n_{\mathrm{M}}}.$$

Let us write the coordination numbers of the metals taking into account that oxygen may be bonded to the metal atoms both of the oxide and the nanoparticle:

$$N^{(\mathrm{NP})}_{\mathrm{M\text{-}M}} = \frac{n^{(\mathrm{NP})}_{\mathrm{M\text{-}M}}}{n^{(\mathrm{NP})}_{\mathrm{M}}}, \quad N^{(\mathrm{Ox})}_{\mathrm{M\text{-}O}} = \frac{n^{(\mathrm{Ox})}_{\mathrm{M\text{-}O}}}{n^{(\mathrm{Ox})}_{\mathrm{M}}} \quad N^{(\mathrm{NP})}_{\mathrm{M\text{-}O}} = \frac{n^{(\mathrm{NP})}_{\mathrm{M\text{-}O}}}{n^{(\mathrm{NP})}_{\mathrm{M}}}.$$

The number of the oxygen bonds on the surface of the nanoparticle $n^{(\mathrm{NP})}_{\mathrm{M\text{-}O}}$ depends on the number of the metal atoms on the surface of the nanoparticle and is expected to be a small value for the large nanoparticles, with a small proportion of the surface atoms.

The respective coordination number $N^{(\mathrm{NP})}_{\mathrm{M\text{-}O}}$ shall also be small.

In EXAFS, all the states are averaged by the metal atoms, i.e.

$$N^{\mathrm{EXAFS}}_{\mathrm{M\text{-}M}} = \frac{n^{(\mathrm{NP})}_{\mathrm{M\text{-}M}}}{n^{(\mathrm{NP})}_{\mathrm{M}} + n^{(\mathrm{Ox})}_{\mathrm{M}}}, \quad N^{\mathrm{EXAFS}}_{\mathrm{M\text{-}O}} = \frac{n^{(\mathrm{NP})}_{\mathrm{M\text{-}O}} + n^{(\mathrm{Ox})}_{\mathrm{M\text{-}O}}}{n^{(\mathrm{NP})}_{\mathrm{M}} + n^{(\mathrm{Ox})}_{\mathrm{M}}}.$$

Let us transit from the number of the bonds to the coordination number:

$$N^{\mathrm{EXAFS}}_{\mathrm{M\text{-}M}} = \frac{n^{(\mathrm{NP})}_{\mathrm{M}} \cdot N^{(\mathrm{NP})}_{\mathrm{M\text{-}M}}}{n^{(\mathrm{NP})}_{\mathrm{M}} + n^{(\mathrm{Ox})}_{\mathrm{M}}},$$

$$N^{\mathrm{EXAFS}}_{\mathrm{M\text{-}O}} = \frac{n^{(\mathrm{NP})}_{\mathrm{M}} \cdot N^{(\mathrm{NP})}_{\mathrm{M\text{-}O}} + n^{(\mathrm{Ox})}_{\mathrm{M}} \cdot N^{(\mathrm{Ox})}_{\mathrm{M\text{-}O}}}{n^{(\mathrm{NP})}_{\mathrm{M}} + n^{(\mathrm{Ox})}_{\mathrm{M}}}$$

and designate a ratio of the atom numbers (concentrations): $\xi = \frac{n^{(\mathrm{NP})}_{\mathrm{M}}}{n^{(\mathrm{NP})}_{\mathrm{M}} + n^{(\mathrm{Ox})}_{\mathrm{M}}}$. The magnitude ξ has a meaning of the proportion of the platinum atoms in the composition of the nanoparticle. At this, $1 - \xi = \frac{n^{(\mathrm{Ox})}_{\mathrm{M}}}{n^{(\mathrm{NP})}_{\mathrm{M}} + n^{(\mathrm{Ox})}_{\mathrm{M}}}$ is a proportion of the platinum atoms in the oxide. Then, the coordination numbers obtained from EXAFS:

$$N^{\mathrm{EXAFS}}_{\mathrm{M\text{-}M}} = \xi \cdot N^{(\mathrm{NP})}_{\mathrm{M\text{-}M}} \tag{A1}$$

$$N^{\mathrm{EXAFS}}_{\mathrm{M\text{-}O}} = \xi \cdot N^{(\mathrm{NP})}_{\mathrm{M\text{-}O}} + (1 - \xi) \cdot N^{(\mathrm{Ox})}_{\mathrm{M\text{-}O}}. \tag{A2}$$

For the equality (A2) we find ξ:

$$\xi = \frac{N^{(\mathrm{Ox})}_{\mathrm{M\text{-}O}} - N^{\mathrm{EXAFS}}_{\mathrm{M\text{-}O}}}{N^{(\mathrm{Ox})}_{\mathrm{M\text{-}O}} - N^{(\mathrm{NP})}_{\mathrm{M\text{-}O}}},$$

then, from (A1) we obtain a relation between the coordination numbers of the metal atoms in the nanoparticle and those obtained from EXAFS:

$$N^{\mathrm{EXAFS}}_{\mathrm{M\text{-}M}} = \frac{N^{(\mathrm{Ox})}_{\mathrm{M\text{-}O}} - N^{\mathrm{EXAFS}}_{\mathrm{M\text{-}O}}}{N^{(\mathrm{Ox})}_{\mathrm{M\text{-}O}} - N^{(\mathrm{NP})}_{\mathrm{M\text{-}O}}} N^{(\mathrm{NP})}_{\mathrm{M\text{-}M}}.$$

And, finally, by neglecting oxidation of the atoms on the surface of the nanoparticle $N^{(\mathrm{Ox})}_{\mathrm{M\text{-}O}} \gg N^{(\mathrm{NP})}_{\mathrm{M\text{-}O}}$, we obtain:

$$N^{(\mathrm{NP})}_{\mathrm{M\text{-}M}} = \frac{N^{(\mathrm{Ox})}_{\mathrm{M\text{-}O}}}{N^{(\mathrm{Ox})}_{\mathrm{M\text{-}O}} - N^{\mathrm{EXAFS}}_{\mathrm{M\text{-}O}}} N^{\mathrm{EXAFS}}_{\mathrm{M\text{-}M}}.$$

Appendix 2

Selection of the hyperparameters of the used models

For the CatBoostRegressor model, the following hyperparameters were selected within the limits:

- iterations: [50, 100, 500, 1000, 1500],
- depth: [5, 6, 7, 8, 10, 12],
- learning_rate: [0.001, 0.01, 0.05, 0.1, 0.5],
- l2_leaf_reg: [1, 3, 5, 10].

Using a grid search method, the following values of the hyperparameters were found: iterations = 100, depth = 7, learning_rate = 0.5, l2_leaf_reg = 5.

The hyperparameters of the neuron networks were tuned by applying a combined approach that includes rough selection and step-by-step complication of the architecture. We have started from simple configurations and gradually

increased the model complexity by adding a layer by a layer and correcting other parameters until we observed overlearning features at a small subsample of the coaching data (20 %). It allowed determining an optimal balance between the model complexity and its generalizability.

Besides, various types of the architectures were tested, including the multi-layer perceptron (MLP) and the convolutional neural network (CNN). The comparative analysis has shown that the multi-layer perceptron demonstrated higher efficiency for solving the set problem, which was confirmed by the quality metrics on the validation and test data.