Chapter 7

Machine-Learning Assisted Structure Determination of Metallic Nanoparticles: A Benchmark

Yuewei Lin[*,¶], Mehmet Topsakal[*,†], Janis Timoshenko[‡], Deyu Lu[†], Shinjae Yoo[*] and Anatoly I. Frenkel[‡,§]

[*]*Computational Science Initiative, Brookhaven National Laboratory, Upton, NY 11973, USA*
[†]*Center for Functional Nanomaterials, Brookhaven National Laboratory, Upton, NY 11973, USA*
[‡]*Department of Materials Science and Chemical Engineering, Stony Brook University, NY 11790, USA*
[§]*Division of Chemistry, Brookhaven National Laboratory, Upton, NY 11973, USA*
[¶]*ywlin@bnl.gov*

X-ray absorption spectroscopy (XAS), which carries rich structural and chemical information of the sample, is a widely used experimental technique for material characterization in diverse scientific fields. This chapter is devoted to the machine learning-based determination of three-dimensional structures of metallic nanoparticles from their spectra. Once built, the machine learning models can be used for parsing through a large volume of experimental spectra in a short time, thus enabling on-the-fly analysis of high-throughput and rapid data collection measurements. As a benchmark test, we compared three regression models, i.e., Gradient boosted trees, shallow/deep multilayer perceptron, and one-dimensional convolutional neural networks. The results showed that the neural networks usually performed better than most tree-based models, while the deep models tended to exhibit higher performance than the shallow ones. Considering the difference between training and testing data, we also evaluated transfer learning, and showed that a significant performance increase can be achieved with the help of partially labeled training data in the target domain. Finally, we demonstrated the high potential of machine learning-based approaches in applications for material science.

Chapter 7: Machine-Learning Assisted Structure Determination of Metallic Nanoparticles: A Benchmark
https://www.worldscientific.com/doi/10.1142/9789811204579_0007

1. Introduction

X-ray absorption spectroscopy (XAS) is a widely used experimental technique for materials characterization in diverse scientific fields including condensed matter physics, materials science, chemistry, earth science, and biology.[1,2] X-rays have sufficient energy to eject a core electron from an atom to the material's empty states and continuum, giving rise to X-ray absorption near edge structure (XANES) and extended X-ray absorption fine structure (EXAFS), respectively, portions of the X-ray absorption coefficient. Transitions from core to final states are determined by well-defined quantum mechanical selection rules, and each element has specific binding energy. Therefore, X-ray absorption spectra are element-specific and carry rich structural and chemical information of the sample.

Synchrotron-based XAS is the state-of-the-art of X-ray techniques, which takes the advantage of intense and tunable X-ray beams at radiation sources. An emerging research frontier is to use synchrotron-based XAS to interrogate functional materials under operando (latin for working/operating) conditions in order to gain insights into the underlying mechanisms in complex chemical and electrochemical processes. In this context, it is essential to decipher local structural information from spectra measured at various stages of a process such as chemical reaction or phase transformation. For example, one wants to detect possible growth of a metal nanocatalyst and control it in real time, during chemical reaction, to prevent catalyst deactivation. So a major challenge is to solve the inverse problem in materials characterization, i.e. determining key local structural motifs from the XAS spectra in real time. While EXAFS is commonly used for solving the unknown local structure around X-ray absorbing species, XANES has unique advantages for structural refinement *in operando* studies.[3] XANES is sensitive to local electronic structure and local point group symmetry of the absorbing site. It has higher tolerance than EXAFS in structural inhomogeneity, and is less sensitive to disorder effects than EXAFS. Finally, XANES can be acquired at harsher reaction conditions and with better time resolution than most of EXAFS data.

The access to these large facilities is quite limited — with over 50 synchrotron light sources around the world[4] — due to the high construction cost and operating expenses. Users often experience long waiting time before performing their experiments in a very short time window, ranging from several hours to a couple of days at the beam line. Due to these constraints, it is timely to revisit the ways through which useful information is extracted

Chapter 7: Machine-Learning Assisted Structure Determination of Metallic Nanoparticles: A Benchmark
https://www.worldscientific.com/doi/10.1142/9789811204579_0007

from the vast amount of XANES data. On the other hand, it is also crucial to make fast decisions to guide experimental processes, especially during *in situ* operations, such as nanoparticle growth or catalytic reactions. Machine-learning approaches applied to interpretation of XANES may offer solutions for these challenges.

To illustrate what new opportunities for nanomaterials characterization "on-the-fly" and reactions "on demand" emerge with the advent of new data processing and analysis methods, we describe here one case study. It is devoted to the machine-learning determination of three-dimensional structure of metallic nanoparticles from their XANES spectra. In this work, we apply supervised machine learning approaches to find the hidden relationship between the XANES spectra and the descriptors of nanoparticle structure. Once built, the machine learning models can be used for parsing through a large volume of experimental spectra in a short time, thus enabling on-the-fly analysis of high-throughput and rapid data collection measurements. An immediate challenge in this approach is the availability of a large representative, labeled training dataset with thousands of data points. Clearly, it would be impractical to attempt to construct such dataset from experimental measurements, because there is only a limited number of such unique spectra for each material. Here we overcame this bottleneck by constructing the training set via *ab initio* XANES simulations validated against experiment. By using theoretical simulations, we can generate a large number of spectra, corresponding to well-defined structure motifs.

In our approach, we use average coordination numbers (CNs) for the first few coordination shells $\{C_1,C_2,C_3,\ldots\}$ that are known to characterize the size and 3D shape of a nanoparticle with close-packed or nearly close-packed structure.[5] Next, we construct a training dataset using *ab initio* codes FEFF[6] and FDMNES.[7] We generate theoretical XANES $x^i(E)$ (here E is X-ray photon energy) for nanoparticles of different sizes/shapes. The sets of corresponding average CNs can be calculated as $C_j = \sum_i n_{ij}/N$, where N is the total number of atoms in the particle, and n_{ij} is the total number of atoms in the jth coordination shell of the ith atom in the nanoparticle. The machine learning models are then defined as a nonlinear function $\mathrm{h}(x^i, \overrightarrow{\Theta}) \rightarrow \{\tilde{C}_1,\tilde{C}_2,\tilde{C}_3,\ldots\}^i$ that uses as input a preprocessed and discretized XANES spectrum x^i and returns a vector $\{\tilde{C}_1,\tilde{C}_2,\tilde{C}_3,\ldots\}^i$. During the training process, we fit the NN parameters so that the distance between the true CNs vector $\{C_1,C_2,C_3,\ldots\}$ and NN output vector $\{\tilde{C}_1,\tilde{C}_2,\tilde{C}_3,\ldots\}$ is minimized for all spectra in our training set. Knowing the CNs, one can then proceed

Chapter 7: Machine-Learning Assisted Structure Determination of Metallic Nanoparticles: A Benchmark
https://www.worldscientific.com/doi/10.1142/9789811204579_0007

to estimate the corresponding NPs size and shape, following the established prescription.[5] For validation, we use particle-averaged XANES data for particles that were used to construct training dataset as well as for particles of other shapes and sizes with fcc-type structure, truncated by (100) and (111) planes, and also with icosahedral and hcp structures. Further details on how training and validation data were generated can be found in Ref. 8.

As the goal is to predict continuous, numeric variables, the coordinate number prediction problem is a typical regression task. In this work, we evaluate three major powerful and widely used regression models, i.e., the gradient boosted trees, multilayer perceptron (MLP) and one-dimensional convolutional neural networks (1D-CNN).

2. Regression Methods

In this section, we will briefly introduce and evaluate three regression models used in this work.

Gradient boosted trees (GBT). Gradient boosted trees (GBT) is an efficient machine learning model that ensembles a set of decision trees.[9] Unlike the bagging-based method, e.g., random forest, which could parallelly train each tree, the GBT computes a sequence of simple trees, where each successive tree is trained for the prediction residuals of the preceding tree. This way, with each tree built, the model becomes more expressive.

In this work, we train one GBT model for each coordination number, so we have four models in total. In each model, we train 100 simple trees with the depth of three.

Multilayer perceptron (MLP). A perceptron (neuron) is a functional block that could be a precursor to many modern larger neural networks. As shown in the left subfigure in Figure 1, a typical perceptron is simple computational units that have weighted input signals and produce an output signal using an (nonlinear) activation function.[10]

The MLP is a network that arranged by a number of perceptron. A typical MLP is shown in the right subfigure in Figure 1. A column of perceptrons is called a layer and MLP could be consists of multiple layers. In MLP, a layer is fully connected with its neighbor layers. The leftmost and rightmost layers are the input layer and output layer, respectively. Layers between them are called hidden layers because that are not directly exposed to the input.

We evaluate two structures of MLP in this work, one is the shallow structure which includes two layers with 400 nodes, and the other one is relatively deep structure which has five layers with 400, 400, 200, 200, 100

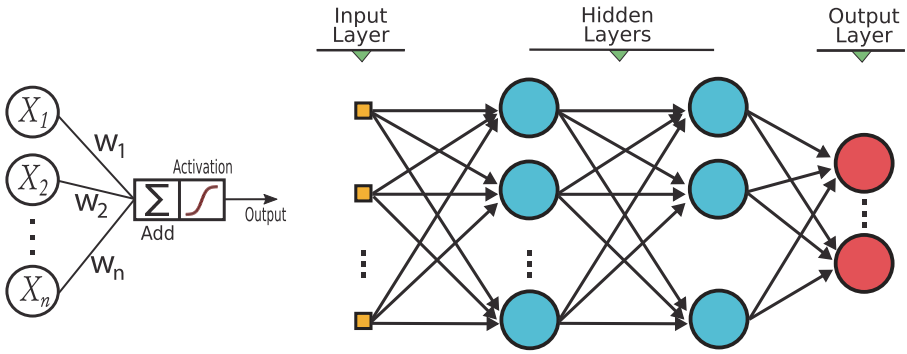

Figure 1. An illustration of an MLP structure.

nodes, respectively. In each layer, the tanh function is used as the activation function.

One-dimensional convolutional neural networks (1D-CNN). We also test one-dimensional convolutional neural networks (1D-CNN), which have shown consistently excellent performance in many applications, such as audio recognition, natural language processing, and etc.[11] 1D-CNN can be considered as a special case of MLP with local connection and shared weights. This way, it has much less number of weights compared with MLP. It also has great capability of extracting local features and the receptive field of the local feature extractors could be hierarchically extended from lower layers to higher layers.

In our work, we build the 1D-CNN which have two convolution layers with 32 and 64 kernels, respectively, and two fully connected layers. The filter size is 30. In each convolutional layer, the rectified linear unit (ReLU) is used as the activation function. Note that we do not utilize any pooling.

In Figure 2, the scatter plots qualitatively show the comparison of the prediction values obtained from four different regression models and the ground truth, with the mean absolute errors (MAE) for four different coordination numbers.

3. Transfer Learning

Most machine learning methods often assume that the training data and testing data are from the same feature space and follow similar distributions. However, this assumption may not be true in many real applications. Namely, the training data are obtained from one domain that we call source domain, while the testing data come from a different domain that we call target

Chapter 7: Machine-Learning Assisted Structure Determination of Metallic Nanoparticles: A Benchmark
https://www.worldscientific.com/doi/10.1142/9789811204579_0007

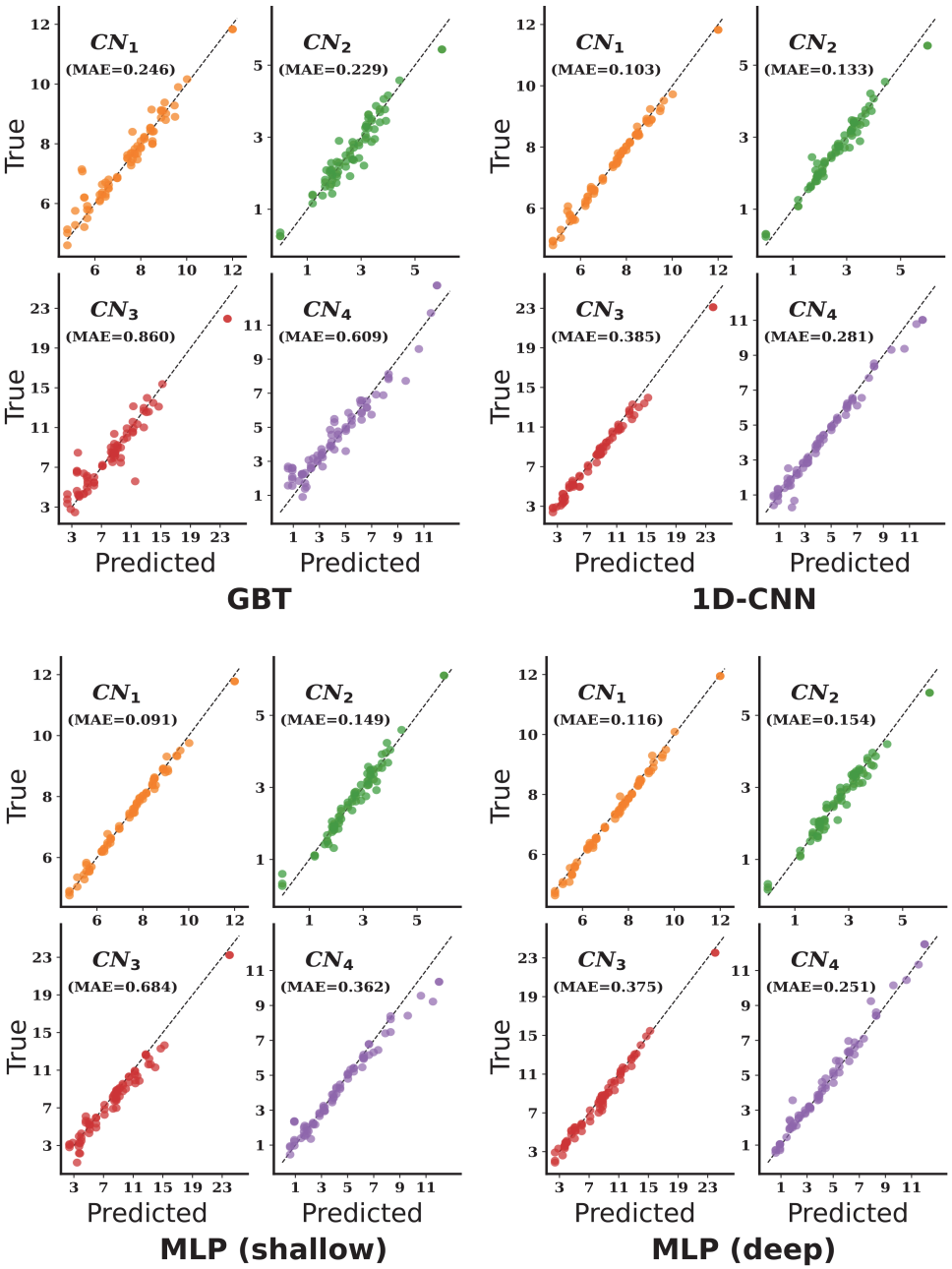

Figure 2. Results of regression methods without using transfer learning.

https://www.worldscientific.com/doi/10.1142/9789811204579_0007

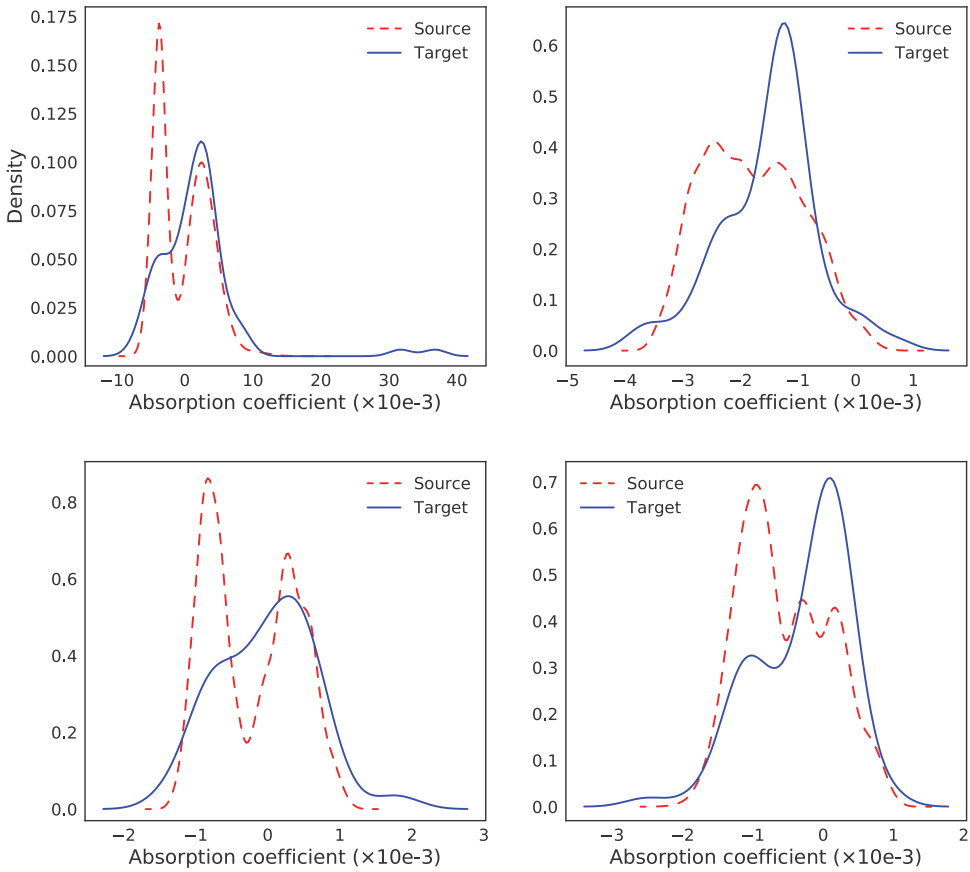

Figure 3. Illustrations of the distribution shift between source and target data.

domain. In this case, the data samples show different distributions in the feature space. In Figure 3, we show several examples that illustrate the difference distributions between the source and target domains.

These domain differences lead to a dilemma that[12]: (1) directly applying the models trained from one domain to another may result in significant degraded performance and (2) labeling large number of data in each domain as training samples would be very expensive. The dilemma consequently poses the transfer learning opportunity, namely how to utilize the information in a source domain to the target domain.

Recently, the works in machine learning and computer vision areas usually apply the transfer learning as the following two steps[13]: (1) a pre-training

https://www.worldscientific.com/doi/10.1142/9789811204579_0007

step, training a model on a source domain, which usually has a large number of labeled data, and (2) a fine tuning step, treating the pre-trained model as the initialized model, rather than randomly initialized, and training on labeled data from target domain. In this work, we also apply this strategy. Specifically, we first train our model on the source data using the deep MLP, and use it as the initialized model to train on target data. Moreover, it is well known that the deep neural networks could be considered as the hierarchal feature extractors, as the lower layers tend to extract the low level features, while the higher layers tend to extract the high level task specific features. Intuitively, for two related data, very similar or even the same low level extractors should be used. Therefore, it is straightforward to freeze the lower layers' weights and only learn the weights in higher layers. Specifically, in this work, we freeze the first two layers and only update the weights in three higher layers. This way, we can also reduce the over fitting, considering the amount of target training data is pretty limited.

We evaluate transfer learning strategy with different numbers (20%, 30%, 40% and 50%) of target training data are used. In each setting, we randomly select the target training data for fine-tuning, and use this fine-tuned model for predicting the rest of target data. We present qualitative results in Figure 4 and quantitative results in Table 1.

4. Discussions

In this work, we compared three regression models, i.e., gradient boosted trees (GBT), shallow (deep) MLP and 1D-CNN. In Figure 2, GBT shows the lowest performance, shallow MLP shows the medium performance, while 1D-CNN and deep MLP reach the best. It follows the general trend observed in many machine learning applications. With the large number of weights and the hierarchical structure, neural networks usually showed the better performance than most tree based models. In the meantime, the deep models tend to exhibit higher performance than the shallow ones, if a large number of training data are available.

Considering the difference between training and testing data, we also evaluated the transfer learning. With the help of partial labeled training data in the target, the performance increases significantly. Even though we only had limited number of labeled training data in the target set, up to 36 samples, the mean absolute errors (MAE) drop from 0.2241 (without transfer learning) to 0.1583 (36 labeled samples).

Chapter 7: Machine-Learning Assisted Structure Determination of Metallic Nanoparticles: A Benchmark
https://www.worldscientific.com/doi/10.1142/9789811204579_0007

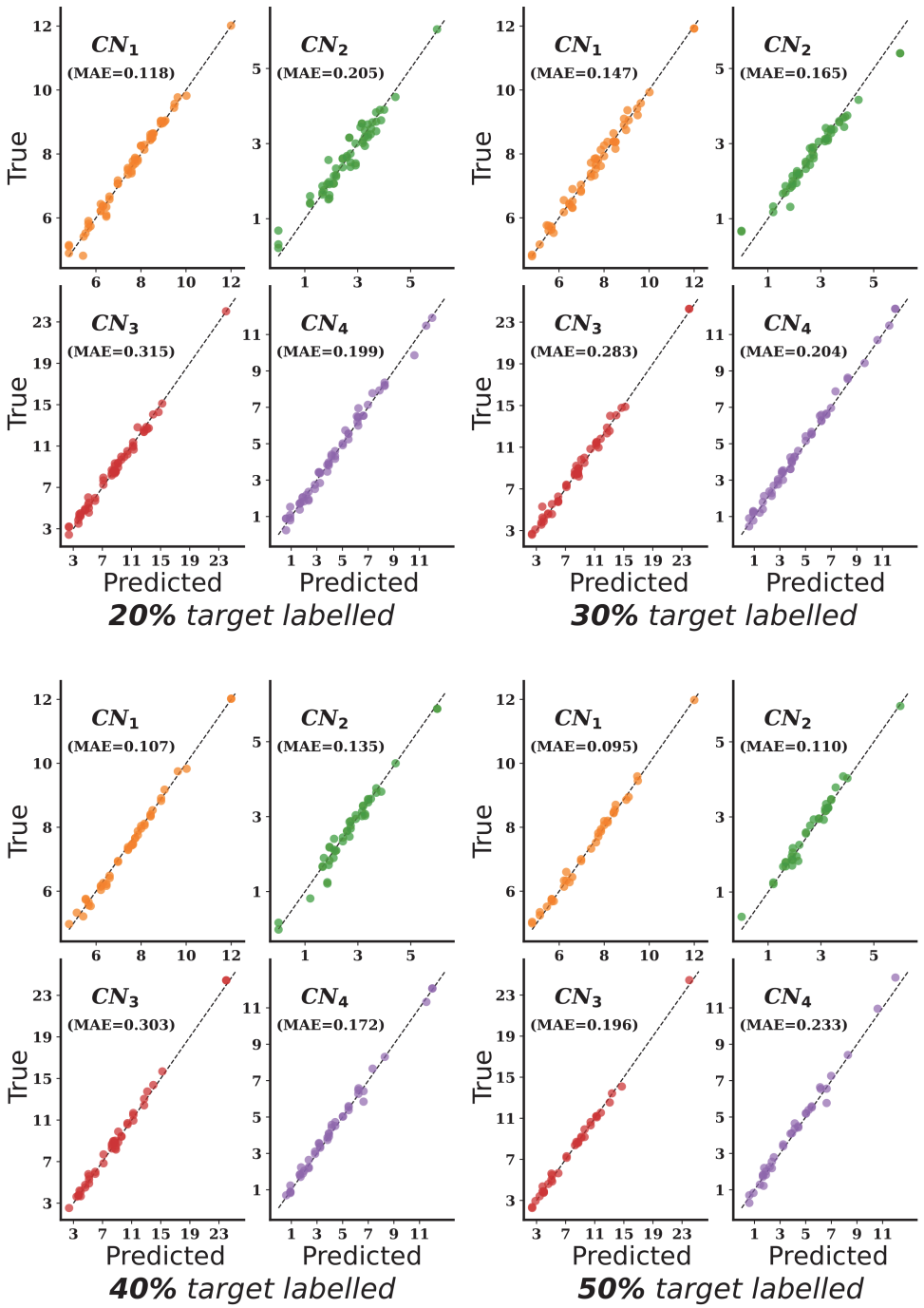

Figure 4. Results of transfer learning.

Chapter 7: Machine-Learning Assisted Structure Determination of Metallic Nanoparticles: A Benchmark
https://www.worldscientific.com/doi/10.1142/9789811204579_0007

Table 1. Results of all comparisons in terms of mean absolute error.

Models		CN_1	CN_2	CN_3	CN_4	Average
Without TL	GBT	0.2457	0.2289	0.8599	0.6087	0.4858
	1D-CNN	0.1029	0.1326	0.3849	0.2809	0.2253
	MLP (shallow)	0.0907	0.1494	0.6838	0.3623	0.3215
	MLP (deep)	0.1163	0.1541	0.3747	0.2512	0.2241
MLP with TL	20% labeled	0.1182	0.2052	0.3152	0.1994	0.2095
	30% labeled	0.1475	0.1654	0.2827	0.2036	0.1998
	40% labeled	0.1066	0.1346	0.3029	0.1720	0.1790
	50% labeled	0.0949	0.1100	0.1956	0.2329	0.1584

5. Examples of Applications

So far in this chapter we have focused on the analysis of the performance of our approach, when it is applied to theoretical data. The usefulness of this method, however, is best demonstrated, when it is applied to the analysis of real experimental XAS data. One needs to be aware that experimental XAS data and theoretically simulated spectra may have differences due to experimental noise, background contribution and systematic artifacts in the experimental data pre-processing as well as due to systematic inaccuracies of the approximations involved in the ab-initio calculations of XANES spectra. Therefore the accuracy in the determination of structure parameters by this method from experimental XANES data can be lower than that demonstrated above in the tests with simulated data.

Nevertheless, in our previous works we showed that the accuracy of the presented approach and its variations is sufficient to extract valuable structural information that, on one hand, can be verified independently for model systems, where complimentary experimental information (e.g., EXAFS data) is available, and, on the other hand, can provide unique insights into the structure of metallic nanoparticles in the cases, when such complimentary data cannot be collected or analyzed. For example, in our study,[8] where this approach was applied for the first time for the analysis of experimental XANES data, we have applied it to the interpretation of Pt L_3-edge XANES data in Pt model nanocatalysts with narrow size- and shape-distributions, supported on γ-Al_2O_3. Using MLP approach, we extracted coordination numbers for the first four coordination shells, and were able to link these CNs to possible particle sizes and shapes (using existing approach[5]). In particular, we have found differences in the shapes for particles, prepared via two different methods, which were confirmed

https://www.worldscientific.com/doi/10.1142/9789811204579_0007

with EXAFS data and microscopy data that were also available for this model system. Since the catalytic activity of nanoparticles depend on particle shape,[14] this information may be important for the design of better catalysts.

In the follow-up studies, we extended this approach to Ag K-edge XANES data and applied it for the first time to the analysis of *in situ* XANES data: investigations of aggregation of silver clusters in ionic liquids.[15,16] This system is important for plasmonic applications. In this case, analysis of XANES data allowed us, first, to independently obtain values of the first shell coordination numbers (and, hence, particle sizes) and use them to verify the CNs, extracted from the conventional EXAFS fitting. Such validation with XANES data was important in this *in situ* study, because poor quality of EXAFS data (due to low concentration of absorbing atoms) limited significantly the accuracy of EXAFS method. Next, using our XANES-based approach, we were able to extract also CNs for the second and third coordination shells, and to use this information to determine the shape of silver particles, which is a crucial parameter for understanding the plasmonic effects. In this case, we have found that the aggregation of silver clusters takes place without coalescence and that the shape and structure of individual clusters within the aggregate is preserved.

In the next study,[17] we focused on the analysis of Cu K-edge XANES data in ultra-dispersed mass-selected copper clusters, prepared in the gas phase and soft-landed on oxide support. XANES data were collected in grazing incidence mode, and acquisition of EXAFS data is not possible for this system due to low sample concentration, alignment issues and Bragg scattering from the support. Extremely small sizes of analyzed clusters (just a few atoms) make this system particularly challenging for investigations with other experimental techniques as well. Moreover, in this material we expect significant deviations in particle structure from that in the bulk. In particular, shortening of Cu–Cu interatomic distance upon reduction of particle size was suggested in the literature,[18] but experimental evidences of this trend were contradictory. To account for this effect, we included one additional degree of freedom in the theoretical data used for MLP training — an effective nearest-neighbor distance — and one corresponding additional node in the MLP output layer. The modified MLP now allowed simultaneous determination of coordination numbers (which can be directly linked to particle sizes) and interatomic distances, and allowed us to confirm the shortening of the latter ones for Cu particles of subnanometer size. Using the information on particle sizes from XANES data, we were able to follow the support-dependent *in situ* agglomeration of Cu clusters during the CO_2 conversion reaction.

Chapter 7: Machine-Learning Assisted Structure Determination of Metallic Nanoparticles: A Benchmark
https://www.worldscientific.com/doi/10.1142/9789811204579_0007

These few examples illustrate the area of possible applications of our method, including *in situ* studies of nanoparticle structure, studies of diluted materials and materials in complex sample environments and/or on strongly attenuating support. We envision that this approach will also be indispensable for studies, where large series of spectra are generated and need to be systematically processed, such as high-throughput studies, time-resolved and/or spatially resolved studies of materials structure: note that after the training of machine learning routine is completed, unlimited number of experimental spectra can be processed within seconds. Finally, similar approaches can be extended for interpretation of not only XANES data, but other spectroscopic data as well. For example, we have already demonstrated that similar ideas, as employed here for the analysis of XANES data, can advance the analysis of EXAFS data as well.[19,20] Together with existing studies, where machine learning is used to assist in theoretical analysis of structure-properties relationship in functional materials,[21–24] these examples demonstrate the high potential of data-science based approaches in materials science applications.

Acknowledgments

Y.L. gratefully acknowledges the support by BNL LDRD 16-039 and BNL LDRD 18-009. A.I.F.'s work was funded by the Division of Chemical Sciences, Geosciences, and Biosciences, Office of Basic Energy Sciences of the US Department of Energy through Grant DE-FG02-03ER15476. D.L. and M.T. used resources of the Center for Functional Nanomaterials, which is a US DOE Office of Science Facility, at Brookhaven National Laboratory under Contract No. DESC0012704. M.T.'s work was supported by LDRD Project at BNL (No. 16-039). RMC-EXAFS simulations were performed on the LASC cluster-type computer at Institute of Solid State Physics of the University of Latvia. X-ray absorption spectra for Pt nanoparticles used in this work were obtained at the beamline 33BM-B of Advanced Photon Source at Argonne National Laboratory and beamline X18B of National Synchrotron Light Source of Brookhaven National Laboratory.

References

1. Koningsberger, D. & Prins, R. *X-ray Absorption: Principles, Applications, Techniques of EXAFS, SEXAFS and XANES*. Chemical Analysis: A Series of Monographs on Analytical Chemistry and Its Applications (Wiley, 1988).
2. van Bokhoven, J. & Lamberti, C. *X-ray Absorption and X-ray Emission Spectroscopy: Theory and Applications*, Vol. 1 (Wiley, 2016).

3. Timoshenko, J., Shivhare, A., Scott, R. W. J., Lu, D. & Frenkel, A. I. Solving local structure around dopants in metal nanoparticles with *ab initio* modeling of X-ray absorption near edge structure. *Phys. Chem. Chem. Phys.* **18**, 19621–19630 (2016).
4. https://lightsources.org/.
5. Glasner, D. & Frenkel, A. I. Geometrical characteristics of regular polyhedra: application to exafs studies of nanoclusters. *AIP Conf. Proce.* **882**(1), 746–748 (2007).
6. Rehr, J. J., Kas, J. J., Vila, F. D., Prange, M. P. & Jorissen, K. Parameter-free calculations of X-ray spectra with feff9. *Phys. Chem. Chem. Phys.* **12**, 5503–5513 (2010).
7. Bunău, O. & Joly, Y. Self-consistent aspects of X-ray absorption calculations. *J. Phys: Conden. Matter* **21**(34), 345501 (2009).
8. Timoshenko, J., Lu, D., Lin, Y. & Frenkel, A. I. Supervised machine-learning-based determination of three-dimensional structure of metallic nanoparticles. *J. Phys. Chem. Lett.* **8**(20), 5091–5098 (2017).
9. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, New York, 2016).
10. Hayakin, S. Neural networks: a comprehensive foundation (1998).
11. Gehring, J., Auli, M., Grangier, D., Yarats, D. & Dauphin, Y. N. Convolutional sequence to sequence learning, preprint (2017), arXiv:1705.03122.
12. Lin, Y., Chen, J., Cao, Y., Zhou, Y., Zhang, L., Tang, Y. Y. & Wang, S. Cross-domain recognition by identifying joint subspaces of source domain and target domain. *IEEE Trans. Cybernetics* **47**(4), 1090–1101 (2017).
13. Girshick, R., Donahue, J., Darrell, T. & Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition* (2014).
14. Mostafa, S., Behafarid, F., Croy, J. R., Ono, L. K., Li, L., Yang, J. C., Frenkel, A. I. & Cuenya, B. R. Shape-dependent catalytic properties of pt nanoparticles. *J. Amer. Chem. Soc.* **132**(44), 15714–15719 (2010).
15. Roese, S., Kononov, A., Timoshenko, J., Frenkel, A. I. & Hövel, H. Cluster assemblies produced by aggregation of preformed ag clusters in ionic liquids. *Langmuir* **34**(16), 4811–4819 (2018).
16. Timoshenko, J., Roese, S., Hövel, H. & Frenkel, A. I. Silver clusters shape determination from *in situ* XANES data. *Radiation Phys. Chem.* (2018).
17. Timoshenko, J., Halder, A., Yang, B., Seifert, S., Pellin, M. J., Vajda, S. & Frenkel, A. I. Subnanometer substructures in nanoassemblies formed from clusters under a reactive atmosphere revealed using machine learning. *J. Phys. Chem. C.* **122**(37), 21686–21693 (2018).
18. Montano, P., Shenoy, G., Alp, E., Schulze, W. & Urban, J. Structure of copper microclusters isolated in solid argon. *Phys. Rev. Lett.* **56**(19), 2076 (1986).
19. Timoshenko, J., Anspoks, A., Cintins, A., Kuzmin, A., Purans, J. & Frenkel, A. I. Neural network approach for characterizing structural transformations by X-ray absorption fine structure spectroscopy. *Phys. Rev. Lett.* **120**(22), 225502 (2018).
20. Timoshenko, J., Wrasman, C. J., Luneau, M., Shirman, T., Cargnello, M., Bare, S. R., Aizenberg, J., Friend, C. M. & Frenkel, A. I. Probing atomic distributions in mono-and bimetallic nanoparticles by supervised machine learning. *Nano Lett.* **19**(1), 520–529 (2018).
21. Graser, J., Kauwe, S. K. & Sparks, T. D. Machine learning and energy minimization approaches for crystal structure predictions: a review and new horizons. *Chem. Mater.* **30**(11), 3601–3612 (2018).

Chapter 7: Machine-Learning Assisted Structure Determination of Metallic Nanoparticles: A Benchmark
https://www.worldscientific.com/doi/10.1142/9789811204579_0007

22. Medford, A. J. Kunz, M. R. Ewing, S. M., Borders, T. & Fushimi, R. Extracting knowledge from data through catalysis informatics. *ACS Catalysis* **8**(8), 7403–7429 (2018).
23. Kitchin, J. R. Machine learning in catalysis. *Nature Catalysis* **1**(4), 230 (2018).
24. Takahashi, K., Takahashi, L., Miyazato, I., Fujima, J., Tanaka, Y., Uno, T., Satoh, H., Ohno, K., Nishida, M., Hirai, K., Ohyama, J., Nguyen, T. N., Nishimura, S. & Taniike, T. The rise of catalyst informatics: towards catalyst genomics. *ChemCatChem* (2019).