Journal of Electron Spectroscopy and Related Phenomena 277 (2024) 147498

Contents lists available at ScienceDirect

Journal of Electron Spectroscopy and Related Phenomena

journal homepage: www.elsevier.com/locate/elspec

Encoder–decoder neural networks in interpretation of X-ray spectra

Jalmari Passilahti, Anton Vladyka, Johannes Niskanen *

Department of Physics and Astronomy, University of Turku, Turun yliopisto, Vesilinnantie 5, 20014, Finland

ARTICLE INFO

Keywords:
X-ray spectroscopy
Machine learning
Neural networks
Encoder–decoder

ABSTRACT

Encoder–decoder neural networks (EDNN) condense information most relevant to the output of the feedforward network to activation values at a bottleneck layer. We study the use of this architecture in emulation and interpretation of simulated X-ray spectroscopic data with the aim to identify key structural characteristics for the spectra, previously studied using emulator-based component analysis (ECA). We find an EDNN to outperform ECA in covered target variable variance, but also discover complications in interpreting the latent variables in physical terms. As a compromise of the benefits of these two approaches, we develop a network where the linear projection of ECA is used, thus maintaining the beneficial characteristics of vector expansion from the latent variables for their interpretation. These results underline the necessity of information recovery after its condensation and identification of decisive structural degrees of freedom for the output spectra for a justified interpretation.

1. Introduction

X-ray spectroscopic methods are sensitive to the local atomistic structure of matter, and are therefore used for characterization [1–5]. In systems with significant structural variation, that of the spectra of individual structures can also be expected. The relationship between a structure and a spectrum is, however, far from trivial due to the origin of the spectral effects in quantum physics of the electronic-nuclear system. Moreover, this relationship may be indicative of only certain structural characteristics for a given spectroscopic method [6]. Thus interpretation of spectra may require an analysis based on a wide set of physically feasible structures, for example from (*ab initio*) molecular dynamics (AI)MD. The formalism used for structural variation is universal from free molecules to liquids and solids. When structural (and spectral) variation is notable, statistical simulation of the ensemble average corresponds to the experiment [7].

To aid these studies machine learning (ML) may be utilized. Models such as neural networks (NN) are known for their ability to make predictions for complicated functions with little computational effort, provided that sufficient training data is available. While still notable, training with standard model selection requires less computational effort than simulations for X-ray spectra. We have recently used these benefits of NNs in an emulator-based component analysis (ECA), a member of project-pursuit algorithms, to identify the structural characteristics a given spectrum is reflective of [6,8–10].

An encoder–decoder neural network (EDNN), illustrated in Fig. 1, is a pipeline of the encoder part, a narrow hidden layer (bottleneck), and the decoder part. If accurate emulation is achieved by such a feed-forward network, knowledge most relevant for spectrum prediction is necessarily condensed in the activation values of the bottleneck neurons. This ability of EDNNs to compress information into few values in the latent space is widely used in computer vision [11], for feature extraction [12] or denoising [13,14]. The aforementioned information condensing takes place intrinsically to training and model selection of ML, and thus knowledge discovery would only require interpretation of the activation values output by the encoder part, passed to the decoder part as an input. As intriguing the potential for automatized information compression by an EDNN sounds for interpretation of X-ray spectra, as relevant it is to find out whether this can be done in practice — and at which cost.

In this work we investigate the structure–spectrum relationship by ML on data of $\sim 10^4$ structures from AIMD and subsequent spectrum calculations for two systems manifesting variation from one local snapshot to another. We benchmark the use of EDNN in the analysis of structural dependence of X-ray spectra using data of the previously studied H_2O molecule [6] and amorphous germanium dioxide [8]. To achieve a meaningful comparison, we closely follow the workflow of these prior analyses. We find that the more flexible EDNN approach captures more target variable variance than the ECA approach of the same number of latent variables. However, finding a reasonable approximate inverse solution for the EDNN architecture is highly untrivial, whereas the ECA allows this owing to inherent linear-only operations. As a compromise we develop an architecture to utilize the strengths of both approaches:

* Corresponding author.
E-mail addresses: jjpass@utu.fi (J. Passilahti), johannes.niskanen@utu.fi (J. Niskanen).

https://doi.org/10.1016/j.elspec.2024.147498
Received 19 June 2024; Received in revised form 21 August 2024; Accepted 25 October 2024
Available online 9 November 2024

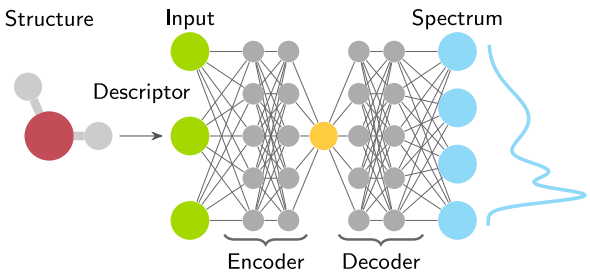

Fig. 1. The principle of spectrum prediction with an encoder–decoder neural network (EDNN). Successful emulation of output for given input requires essential latent information to be condensed into the activation values of the neurons at the bottleneck, which may be only a few for reasonably accurate emulation of the structure–spectrum relationship.

a linear projection encoder (such as ECA), a narrow layer of neurons for latent variables, and a freely adjusting decoder. In this neural-network component analysis (NNCA) approach the basis vectors of ECA are optimized jointly with the training of the subsequent neural network, in the spirit of EDNN, but the constrained architecture allows for stable approximate solution of the inverse problem.

2. Methods

In this work we use two previously studied simulated datasets. The first one consists of 10 000 snapshots from AIMD trajectories for H_2O molecule and respective O K-edge X-ray photoelectron spectra (XPS), X-ray emission spectra (XES) and X-ray absorption spectra (XAS) [6,15], openly available in [16]. The second dataset consists of 13 896 structures and respective Ge Kβ XES of amorphous GeO_2 simulated at 11 different pressures ranging from 0 to 120 GPa. Each data point consists of a local Ge-centered structure based on AIMD simulations [17,18], and the corresponding calculated spectrum [8,18]. For both systems the AIMD used Perdew–Burke–Ernzerhof (PBE) exchange–correlation potential [19]. The molecular H_2O simulation [15] was carried out for the NVE ensemble with the initial kinetic energy equivalent to 10 000 K using Goedecker–Teter–Hutter pseudopotentials [20–22] and triple-ξ TZV2P-MOLOPT-GTH basis set [23], delivered with the used `CP2K` software [24]. The amorphous GeO_2 simulation [17,18] focused on the Γ-point for a supercell of 216 atoms with a plane-wave basis of the cutoff of 395 eV, using the `VIENNA` software package [25].

Successful ML of X-ray spectra requires input data to be presented in a format different to that of the raw data, *i.e.* feature engineering [10]. We encoded the geometry of the H_2O molecule simply in a form of H–O–H angle α, the length b_l of the long O–H bond and the length b_s of the short O–H bond. For amorphous GeO_2 we followed Ref. [8] by describing the local structure as Coulomb matrix [26]. The output features may require engineering as well; whereas spectra tabulated on a grid in energy could be used for the molecular H_2O, the limited amount of data necessitated encoding the spectral information of amorphous GeO_2 into 4 statistical moments (mean, standard deviation, skewness and excess kurtosis) for Kβ'' and Kβ_2 emission lines as done in Ref. [8]. We trained the models using 80% (train set) of the data and evaluated the final performance with the remaining 20% (test set), unseen to the model. Furthermore, we z-score standardized the structural features for the train set to have zero mean and unit variance. For spectra of the H_2O molecule, we used the mean and variance over all channels in the standardization as in Ref. [6]. The spectral moments of amorphous GeO_2 were individually z-score standardized.

We carried out randomized-grid-search model selection using one and two latent variables (bottleneck widths) for both systems. This procedure used 5-fold cross validation, and the best model was trained with the full train set afterwards. In each case, the depth and the width of the network were varied together with the learning rate and the strength of the L_2-regularization. We used the rectified linear unit (ReLU) activation function [27] for these networks. In total we tested over 500 000 hyperparameter combinations: 120 000 for molecular H_2O (20 000 per model type), and 400 000 for the amorphous GeO_2 (200 000 per model type). We used `Python`[28] package `scikit-learn`[29] for these tasks. The model selection search spaces are provided in Supplementary Information.

As a general measure for the performance of models we use the R^2 score (generalized covered target variance):

$$R^2 = 1 - \frac{\mathrm{tr}\left(\tilde{\mathbf{S}}^{\mathrm{T}}\tilde{\mathbf{S}}\right)}{\mathrm{tr}\left(\mathbf{S}^{\mathrm{T}}\mathbf{S}\right)}, \tag{1}$$

where matrix $\mathbf{S}$ contains the true target features (spectra or related measures) as row vectors and $\tilde{\mathbf{S}}$ is the difference between true and corresponding predicted spectral output $\mathbf{S}^{(\mathrm{pred})}$:

$$\tilde{\mathbf{S}} = \mathbf{S} - \mathbf{S}^{(\mathrm{pred})}. \tag{2}$$

The R^2 metric takes values in the range $]-\infty, 1]$, where 1 corresponds to perfect match, and 0 corresponds to the performance of constant prediction of the mean value. The R^2 score is a relative measure with an interpretable range of values independent of unit or absolute scale. Analysis of changes with respect to a baseline spectrum, expected for the chemical entity, is a meaningful task for interpretation of a spectrum, fulfilled by mean-subtracted data and R^2. Last, since structure–spectrum relationship is nonlinear, a small spectral feature may be indicative of a physically interesting structure or a physical mechanism. These points motivate the use of R^2, but we note that any other metric could be used if desired. Additionally, this choice allows for direct comparison to the earlier works by Niskanen et al. [6] and Vladyka et al. [8].

We benchmark EDNN against ECA [6]. This projection-pursuit-spirited analysis utilizes a pre-trained emulator $\mathbf{s}_{\mathrm{emu}}(\mathbf{x})$, such as a NN, to identify relevant input variations in terms of the variations of the output by providing data $\mathbf{S}^{(\mathrm{pred})}$ of Eq. (2) for given structural data points $\mathbf{x}$. The method carries out a decomposition in the input space, in which the basis vectors $\mathbf{v}_j$ are optimized one at a time to maximize the prediction performance score for rank k of the projection

$$\mathbf{x}^{(k)} = \sum_{j=1}^{k} \underbrace{(\mathbf{v}_j \cdot \mathbf{x})}_{=:t_j} \mathbf{v}_j. \tag{3}$$

Here t_j is defined as the inner product of the input vector $\mathbf{x}$ and the jth basis vector $\mathbf{v}_j$. It is inherently assumed that feature-wise z-score-standardized input data points $\mathbf{x}$ are used. Here we use the test dataset to maximize the R^2 score for $\mathbf{s}_{\mathrm{emu}}(\mathbf{x}^{(k)})$ and spectra $\mathbf{s}$ (row vectors of the matrix $\mathbf{S}$) known for structures $\mathbf{x}$. In this iterative procedure the quickness of the ML emulator $\mathbf{s}_{\mathrm{emu}}$ is crucial.

As motivated and discussed in Section 3.3, we implemented an EDNN variant (NNCA, *Neural Network Component Analysis*), where the encoder consisted of a single hidden layer with orthonormality constraints for the basis vectors, without bias and without activation function. This corresponds to the projection step of Eq. (3) for t_j and allows fitting the ECA basis vectors $\mathbf{v}_j$ during the training of the network. In this manner, the model selection phase of ML is also leveraged for ECA. We implemented the aforementioned network with ReLU activation function using `PyTorch`[30] (v. 2.2.1). We ran exhaustive network architecture model selection for the decoder part using the same search spaces that were used for the emulators of the ECA implementations [6,8] which are provided in the Supplementary Information.

3. Results and discussion

Fig. 2 shows the mean spectra of the structural data from AIMD and subsequent spectrum calculations. The gray areas represent the variation in the dataset and in each case the spectra show noticeable

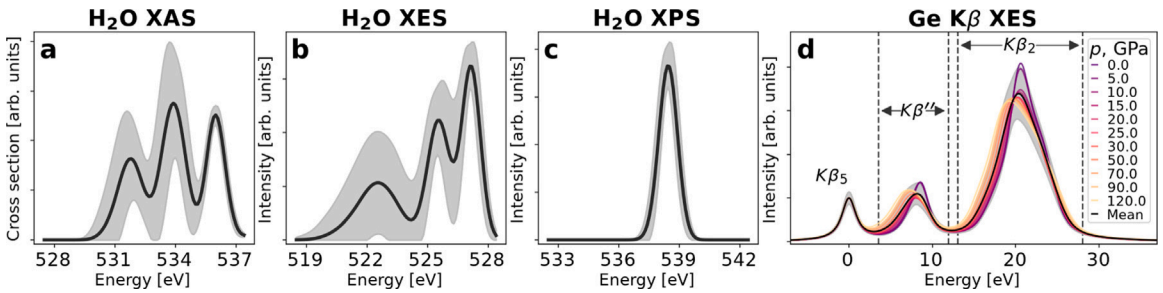

Fig. 2. Spectra for structures from AIMD simulations for (a-c) the H_2O molecule and (d) amorphous GeO_2 at various pressures. Black line depicts the mean and gray areas $\pm\sigma$ cut to zero. For amorphous GeO_2 colored lines represent mean spectra for different pressures and the intervals containing the studied peaks are marked with vertical dashed lines.

Table 1
Generalized covered spectral variance (R^2 score) for molecular H_2O using EDNN, NNCA and ECA with 1 and 2 bottleneck neurons or ECA-components k. Notable improvement is achieved owing to nonlinearity of the encoder. Same train and test splits as in [6] were used with all models.

	k	EDNN	NNCA	ECA [6]
XES	1	0.86	0.79	0.74
	2	1.00	1.00	1.00
XAS	1	0.85	0.76	0.75
	2	0.98	0.92	0.91
XPS	1	1.00	0.99	0.99
	2	1.00	1.00	1.00

structural dependency. Furthermore, in panel (d) a pressure dependency of the Ge Kβ XES spectrum of amorphous GeO_2 is clearly visible. Conveniently, the H_2O molecule has only a few inputs and many outputs while amorphous GeO_2 has the opposite. This contrast may help unveil interesting insights about our solution and the problem in general.

3.1. H_2O molecule

The performance of a dimensionality reduction is measured by the spectral variance that it allows to cover, for which we use the R^2 score. These numbers are presented for ECA and EDNN in Table 1. The table also contains values for NNCA, discussed later in Section 3.3. A clear improvement for test data is seen with the more flexible EDNN, when the bottleneck width equals the corresponding ECA rank. With one-component model for XES, a score improvement of 0.12 units is achieved in comparison to ECA, and the spectral dependence can be completely encoded into two latent variables by using either approach. For XAS, 0.10 units better performance than that of ECA is achieved by one-component EDNN, and the two-component EDNN leaves only a minor 0.02-unit fraction out of the complete spectral behavior. For XPS, a single degree of freedom suffices to describe the whole variation. Using two components, virtually no relevant structural information is lost in the bottleneck in any of the cases.

The structural simplicity of the molecular H_2O system allows for inspection of performance by visualization. As an metric for evaluation of the final model for the H_2O molecule we use the spectral deviation from that at the center of the training set

$$M(\mathbf{x}) := \frac{\|\mathbf{s}_{\text{emu}}(\mathbf{x}) - \mathbf{s}_{\text{emu}}(\mathbf{x}_{\text{cen}})\|_2}{\|\mathbf{s}_{\text{emu}}(\mathbf{x}_{\text{cen}})\|_2}, \tag{4}$$

where $\mathbf{x}_{\text{cen}}$ is located at origin when using standardized data. Isosurface plots for EDNN activation value for the one-component model are compared to isosurfaces of the metric with the polynomial emulator [6] in Fig. 3 . The figure shows qualitatively similar behavior for M and for the bottleneck activation value of the EDNN. This in turn indicates that the overall spectral behavior is encoded in the activation value. We note that only the relative scale of activation value is meaningful. We conclude that the bottleneck activation captures the content of the

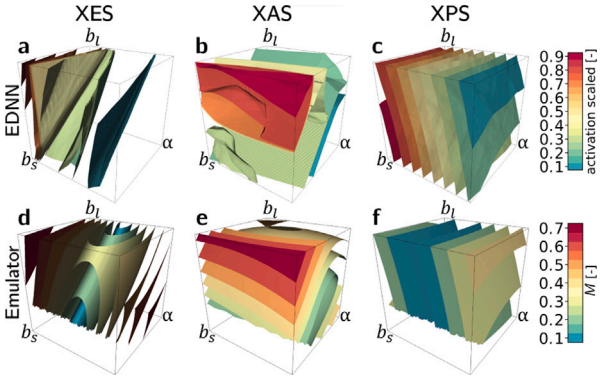

Fig. 3. Isosurface plots (a-c) of the one-component EDNN model activation values compared to (d-f) isosurfaces of metric M defined in Eq. (4) a polynomial model [6] for the spectra of the H_2O molecule. Drawn following Ref. [6]. The H–O–H angle, the length of the long O–H bond, and the length of the short O–H bond, are denoted as α, b_l, b_s, respectively.

difference metric M and spectral change in the structural space. We note that some of the XES spectrum isosurfaces are close to each other, indicating structural regions of high spectral sensitivity, in addition to identification of this spectrally dominant direction in the structural space. Furthermore, the t-score isosurfaces would be planes for ECA due to the linear formalism, and therefore best one-component ($k = 1$) ECA performance is obtained for XPS. The deviation of EDNN from this planar behavior is owing to the flexibility of the respective encoder.

For interpretation of spectra, or latent variables derived thereof, a structural correspondent might be searched for. While the EDNN succeeds in the forward problem with a bottleneck, finding an inverse for the network proves problematic. We trace this to non-bijective nature of the structure–spectrum problem, manifested here by ill-conditioned weight matrices in the network, which will amplify errors when propagation to the inverse direction is carried out. In our try to overcome these problems, we carried out a separate model selection with a bijective activation function and an architecture with as many square weight matrices as possible. The details and results for the H_2O molecule are documented in Supplementary Information, and we will return to this topic with amorphous GeO_2.

3.2. Amorphous GeO_2

An ECA analysis of Ge Kβ X-ray emission spectrum (XES) of amorphous GeO_2 has been previously carried out by Vladyka et al. [8]. In the work prediction of full spectra was not achieved but instead the authors focused on eight statistical moments for two spectral lines as target features. The generalized covered variance (R^2 score) for the features with the EDNN and ECA are given in Table 2, where EDNN is again seen to achieve greater scores than ECA. Notably, the score of the one-component EDNN surpasses that of the corresponding two-component

Table 2
Generalized covered spectral variance (R^2 score) for the spectral moments of the Ge Kβ XES of amorphous GeO_2 using EDNN, NNCA and ECA with one and two bottleneck neurons/ECA-components k. Again significant improvement is observed owing to nonlinearity of the encoder of the EDNN. Same train and test splits as in [8] were used with all models. For details, see text.

	k	EDNN	NNCA	ECA [8]
XES moments	1	0.84	0.84	0.77
	2	0.90	0.88	0.83

ECA model. Thus, more significant portion of the information affecting the spectral moments can be condensed even into a single latent variable using EDNN architecture.

As shown by Vladyka et al. ECA provides an approximate solution to the spectrum-to-structure inverse problem [8]. In the work, scores t_1 and t_2 could be reconstructed from spectral moments, after which general trends in the Coulomb matrix and in the interactomic distances from the active Ge site could be reconstructed. Our trials using a bijective activation function Leaky ReLU [31] for EDNN were in general unsuccessful in reconstructing the bottleneck activation values. As with the H_2O molecule, we applied the specific architecture of maximal number of square weight matrices for this specific task of inverse (see Supplementary Information). The data from the amorphous GeO_2 and molecular H_2O reveal a trend: the larger matrices, whether in encoder or decoder, cause error when an inverse is attempted. In addition, the step from encoder to the bottleneck is always problematic.

3.3. NNCA: ECA implemented during training

For ECA the problem of finding an inverse may not be as severe as for EDNN: while reconstruction of the **t** scores may be susceptible to non-bijective behavior, at least structural interpretation of these scores is straightforward by application of Eq. (3) providing an approximate structural data point [8]. Even though in some cases the activation values of the neurons at the bottleneck could be reconstructable, the EDNN would then require approximate structural reconstruction to achieve the same. We combined the automatized optimization for information condensing by the EDNN to the interpretability brought by ECA into neural network component analysis (NNCA), illustrated in Fig. 4. In this network architecture, the first layer carries out the projection step of ECA to latent variables (without the expansion of Eq. (3)). As the basis vectors $\{\mathbf{v}_j\}$ then are optimized during training, and as the whole process is subject to model selection, more covered spectral variance than for ECA is expected. Moreover, NNCA using built-in orthonormality may give better performance for $k > 1$ components than rank-by-rank proceeding ECA, owing to collaborative adaptation of the vectors subject to optimization. Last, latent variables condensed in the bottleneck can be converted back to structural space by simple vector expansion. As a downside of this approach, the hierarchy of the ECA basis vectors in terms of covered target variance is lost.

Tables 1 and 2 indicate the performance of NNCA to expectedly lie in between those of EDNN and ECA, as NNCA is a special case of EDNN and ECA a special case of NNCA. Notably the NNCA manages to match the performance of EDNN with one-component for the spectral moments of the Ge Kβ XES of amorphous GeO_2. This limited comparison hints that the performance of NNCA is closer to EDNN when the system is more complex and particularly when the structural descriptor is more complex.

We studied structural reconstruction by NNCA following closely the earlier ECA procedure of Vladyka and co-workers [8]. In the cited work reasonable, but not complete, pressure-wise reconstruction of ensemble-mean interatomic distances from the active Ge site for Ge Kβ XES of amorphous GeO_2 was achieved from the eight spectral moments. Motivated by earlier observations of unstable approximate inverse, even due to small discrepancies, in the NNCA procedure the

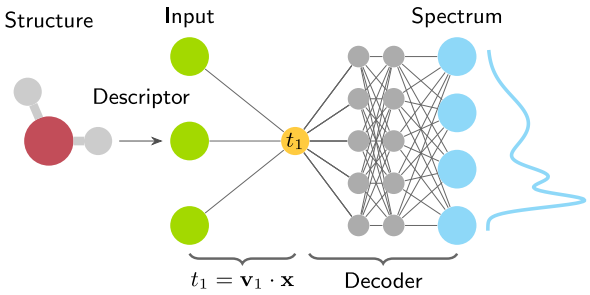

Fig. 4. The principle of implementing ECA as a simple encoder in the NNCA network with k=1. A broader bottleneck implies corresponding orthonormal set of vectors $\{\mathbf{v}_j\}_{j=1}^{k}$ for the projection of the input vector **x**. The matrix operation after input has zero bias followed by linear activation.

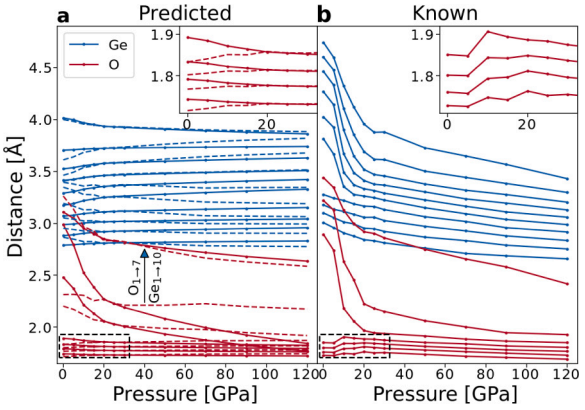

Fig. 5. Mean-structural-parameter-based distances in each pressure from the central Ge atom (a) Reconstructed from the fitted NNCA components for both one and two-component models. Dashed lines correspond to the two-component model. (b) Known mean distances calculated from the atomic coordinates of the test dataset. Drawn following Ref. [8].

activation values **t** of the bottleneck are optimized by least-squares fitting for the target variables, after which the expansion of Eq. (3) is taken. This expansion corresponds to (pseudo-)inverting the first matrix layer between input **x** and scores **t**, illustrated for the NN of Fig. 4.

Fig. 5 depicts the results for each pressure in the dataset together with the known values. The rapid fall of Ge–O interatomic distances up to 20 GPa coincides with the rapid decrease of 4-fold Ge–O coordination in the amorphous GeO_2 [18]. However, the changes related to interatomic Ge–Ge distances are not reconstructed by NNCA. When compared to reconstruction by ECA [8] the NNCA outperforms in regard to Ge–O distances, but leaves out the effect in Ge–Ge distances. Moreover, the one-component model gives the more accurate structural reconstruction out of the two. The applied expansion method predicts the mean value for spectrally irrelevant parameters due to z-score standardization of the input. Thus a criterion probably exists for a component of a basis vector (and the respective input feature) to be deemed indecisive in expansion of Eq. (3). We leave the investigation of this condition for future, because the study requires data on more systems than currently available to us.

4. Conclusions

Recent research has shown that some structural information is irrecoverably lost in a respective X-ray spectrum. Therefore, a justified interpretation of the spectroscopic probe requires identifying the decisive structural degrees of freedom, as the behavior of a spectrum may be dictated by only few structural characteristics. We studied condensing structure–X-ray spectrum relationship by encoder–decoder neural

networks (EDNN) for the H_2O molecule and for amorphous GeO_2. The EDNNs with a bottleneck of k neurons covered more spectral variance than an emulator-based component analysis (ECA) decomposition of the same rank. This happens owing to the flexibility of the encoder and its related adaptivity to nonlinear behavior in the multidimensional input space. However, the improved covered target variance does not come without a cost: interpretation of the latent degrees of freedom behind spectral sensitivity, *i.e.* the activation values at the bottleneck, becomes problematic in itself.

These problems of EDNN could be cured by implementation of the projection of input to latent variables by basis vectors such as in ECA. When this new NNCA approach was implemented as a part of the neural network training and model selection, we observed expected rise in the covered variance to that of ECA. The approach helped to recover the coordination change in amorphous GeO_2 at elevated pressures, in a qualitative agreement with previous studies. However, this improved performance implied the loss of the hierarchy present for the basis vectors, which could in principle be fixed by posterior unitary transformation of the vectors.

Problems in finding inverse functions in the EDNN are indicative of the loss of structural information upon spectrum formation. However this information may be available from the structural simulation data behind the spectrum calculations, which may leak into interpretation of spectra. Due to this risk of overinterpretation of spectra by information leakage, investigation of reconstructable information is an essential aspect of interpreting X-ray spectra with the help of simulations. Indeed, a realistic interpretation of spectra can only comment on structural characteristics that have an effect on them.

CRediT authorship contribution statement

Jalmari Passilahti: Writing – review & editing, Writing – original draft, Visualization, Software, Investigation, Formal analysis. **Anton Vladyka:** Writing – review & editing, Writing – original draft, Validation, Data curation. **Johannes Niskanen:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Johannes Niskanen reports financial support was provided by Academy of Finland. Jalmari Passilahti reports financial support was provided by Academy of Finland. Anton Vladyka reports financial support was provided by Academy of Finland.

Acknowledgments

Academy of Finland is acknowledged for funding via project 331234. The authors acknowledge CSC – IT Center for Science, Finland, and the FGCI - Finnish Grid and Cloud Infrastructure for computational resources. The authors thank Mr. E.A. Eronen for discussions.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.elspec.2024.147498.

Data availability

Data will be made available on request.

References

[1] K. Siegbahn, A. Nordling, R. Fahlman, K. Nordberg, J. Hamrin, G. Hedman, T. Johansson, S.-E. Bergmark, I. Karlsson, L.B. Lindgren, ESCA – Atomic, Molecular and Solid State Structure Studied By Means of Electron Spectroscopy, North-Holland Publishing Company, Amsterdam, 1967.
[2] K. Siegbahn, C. Nordling, G. Johansson, J. Hedman, P.F. Heden, K. Hamrin, U. Gelius, T. Bergmark, L.O. Werme, R. Manne, Y. Baer, ESCA – Applied to Free Molecules, North-Holland Publishing Company, Amsterdam, 1969.
[3] J. Stöhr, NEXAFS Spectroscopy, Springer Verlag, 1992.
[4] W. Schülke, Electron dynamics by inelastic X-ray scattering, vol. 7, Oxford University Press, 2007.
[5] P. Zimmermann, S. Peredkov, P.M. Abdala, S. DeBeer, M. Tromp, C. Müller, J.A. van Bokhoven, Modern X-ray spectroscopy: XAS and XES in the laboratory, Coord. Chem. Rev. 423 (2020) 213466, http://dx.doi.org/10.1016/j.ccr.2020.213466.
[6] J. Niskanen, A. Vladyka, J. Niemi, C. Sahle, Emulator-based decomposition for structural sensitivity of core-level spectra, R. Soc. Open Sci. 9 (6) (2022) 220093, http://dx.doi.org/10.1098/rsos.220093.
[7] M.P. Allen, D.J. Tildesley, Computer Simulation of Liquids, Oxford University Press, Oxford, UK, 1987.
[8] A. Vladyka, C.J. Sahle, J. Niskanen, Towards structural reconstruction from X-ray spectra, Phys. Chem. Chem. Phys. 25 (9) (2023) 6707–6713, http://dx.doi.org/10.1039/d2cp05420e.
[9] E.A. Eronen, A. Vladyka, F. Gerbon, C.J. Sahle, J. Niskanen, Information bottleneck in peptide conformation determination by x-ray absorption spectroscopy, J. Phys. Commun. 8 (2024) 025001, http://dx.doi.org/10.1088/2399-6528/ad1f73.
[10] E.A. Eronen, A. Vladyka, C.J. Sahle, J. Niskanen, Structural descriptors and information extraction from x-ray emission spectra: aqueous sulfuric acid, Phys. Chem. Chem. Phys. (2024) http://dx.doi.org/10.1039/D4CP02454K.
[11] H. Noh, S. Hong, B. Han, Learning deconvolution network for semantic segmentation, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1520–1528.
[12] J. Masci, U. Meier, D. Cireşan, J. Schmidhuber, Stacked convolutional auto-encoders for hierarchical feature extraction, in: Artificial Neural Networks and Machine Learning–ICANN 2011: 21st International Conference on Artificial Neural Networks, Springer, 2011, pp. 52–59.
[13] C. Xing, L. Ma, X. Yang, Stacked denoise autoencoder based feature extraction and classification for hyperspectral images, J. Sensors 2016 (2016) 1–10, http://dx.doi.org/10.1155/2016/3632943.
[14] T. Konstantinova, L. Wiegart, M. Rakitin, A.M. DeGennaro, A.M. Barbour, Noise reduction in x-ray photon correlation spectroscopy with convolutional neural networks encoder–decoder models, Sci. Rep. 11 (1) (2021) http://dx.doi.org/10.1038/s41598-021-93747-y.
[15] J. Niskanen, A. Vladyka, J.A. Kettunen, C.J. Sahle, Machine learning in interpretation of electronic core-level spectra, J. Electron Spectrosc. Relat. Phenom. 260 (2022) 147243, http://dx.doi.org/10.1016/j.elspec.2022.147243.
[16] J. Niskanen, A. Vladyka, J. Niemi, C.J. Sahle, Data from: Emulator-based decomposition for structural sensitivity of core-level spectra, 2022, http://dx.doi.org/10.5061/dryad.dncjsxm1m.
[17] X. Du, J.S. Tse, Oxygen packing fraction and the structure of silicon and germanium oxide glasses, J. Phys. Chem. B 121 (47) (2017) 10726–10732, http://dx.doi.org/10.1021/acs.jpcb.7b09357.
[18] G. Spiekermann, C.J. Sahle, J. Niskanen, K. Gilmore, S. Petitgirard, C. Sternemann, J.S. Tse, M. Murakami, Sensitivity of the Kβ X-ray emission line to coordination changes in GeO2 and TiO2, J. Phys. Chem. Lett. 14 (7) (2023) 1848–1853, http://dx.doi.org/10.1021/acs.jpclett.3c00017.
[19] J.P. Perdew, K. Burke, M. Ernzerhof, Generalized gradient approximation made simple, Phys. Rev. Lett. 77 (1996) 3865–3868, http://dx.doi.org/10.1103/PhysRevLett.77.3865.
[20] S. Goedecker, M. Teter, J. Hutter, Separable dual-space gaussian pseudopotentials, Phys. Rev. B 54 (1996) 1703–1710, http://dx.doi.org/10.1103/PhysRevB.54.1703.
[21] C. Hartwigsen, S. Goedecker, J. Hutter, Relativistic separable dual-space Gaussian pseudopotentials from H to Rn, Phys. Rev. B 58 (1998) 3641–3662, http://dx.doi.org/10.1103/PhysRevB.58.3641.
[22] M. Krack, Pseudopotentials for H to Kr optimized for gradient-corrected exchange–correlation functionals, Theor. Chem. Acc. 114 (1) (2005) 145–152, http://dx.doi.org/10.1007/s00214-005-0655-y.
[23] J. VandeVondele, J. Hutter, Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases, J. Chem. Phys. 127 (11) (2007) 114105, http://dx.doi.org/10.1063/1.2770708.
[24] T.D. Kühne, M. Iannuzzi, M. Del Ben, V.V. Rybkin, P. Seewald, F. Stein, T. Laino, R.Z. Khaliullin, O. Schütt, F. Schiffmann, D. Golze, J. Wilhelm, S. Chulkov, M.H. Bani-Hashemian, V. Weber, U. Borštnik, M. Taillefumier, A.S. Jakobovits, A. Lazzaro, H. Pabst, T. Müller, R. Schade, M. Guidon, S. Andermatt, N. Holmberg, G.K. Schenter, A. Hehn, A. Bussy, F. Belleflamme, G. Tabacchi, A. Glöß, M. Lass, I. Bethune, C.J. Mundy, C. Plessl, M. Watkins, J. VandeVondele, M. Krack, J. Hutter, CP2K: An electronic structure and molecular dynamics software package - Quickstep: Efficient and accurate electronic structure calculations, J. Chem. Phys. 152 (19) (2020) 194103, http://dx.doi.org/10.1063/5.0007045.

[25] G. Kresse, J. Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set, Phys. Rev. B 54 (1996) 11169–11186, http://dx.doi.org/10.1103/PhysRevB.54.11169.

[26] M. Rupp, A. Tkatchenko, K.R. Müller, O.A. Von Lilienfeld, Fast and accurate modeling of molecular atomization energies with machine learning, Phys. Rev. Lett. 108 (5) (2012) 058301, http://dx.doi.org/10.1103/PhysRevLett.108.058301, arXiv:1109.2618.

[27] V. Nair, G.E. Hinton, Rectified linear units improve restricted boltzmann machines, in: Proceedings of the 27th International Conference on Machine Learning, ICML-10, 2010, pp. 807–814, URL https://www.cs.toronto.edu/%7Efritz/absps/reluICML.pdf.

[28] Python Software Foundation, Python software foundation python language reference version 3.10.9, 2024, Available: https://www.python.org/, (cited 22.05.2024).

[29] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine Learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.

[30] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, Pytorch: An imperative style, high-performance deep learning library, 2019, http://dx.doi.org/10.48550/arXiv.1912.01703, arXiv, arXiv:1912.01703.

[31] A.L. Maas, A.Y. Hannun, A.Y. Ng, et al., Rectifier nonlinearities improve neural network acoustic models, in: Proc. icml, vol. 30, 2013, p. 3, URL https://ai.stanford.edu/amaas/papers/relu_hybrid_icml2013_final.pdf.