The Journal of Open Source Software

Chemiscope: interactive structure-property explorer for materials and molecules

Guillaume Fraux[1], **Rose K. Cersonsky**[1], **and Michele Ceriotti**[1]

1 Laboratory of Computational Science and Modeling, IMX, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

DOI: 10.21105/joss.02117

Software

- Review
- Repository
- Archive

Editor: Christopher R. Madan

Reviewers:

- @kblin
- @mkhorton

Submitted: 12 February 2020
Published: 01 July 2020

Summary

The number of materials or molecules that can be created by combining different chemical elements in various proportions and spatial arrangements is enormous. Computational chemistry can be used to generate databases containing billions of potential structures (Ruddigkeit, Deursen, Blum, & Reymond, 2012), and predict some of the associated properties (Montavon et al., 2013; Ramakrishnan, Dral, Rupp, & Lilienfeld, 2014). Unfortunately, the very large number of structures makes exploring such database — to understand structure-property relations or find the *best* structure for a given application — a daunting task. In recent years, multiple molecular *representations* (Bartók, Kondor, & Csányi, 2013; Behler & Parrinello, 2007; Willatt, Musil, & Ceriotti, 2019) have been developed to compute structural similarities between materials or molecules, incorporating physically-relevant information and symmetries. The features associated with these representations can be used for unsupervised machine learning applications, such as clustering or classification of the different structures, and high-throughput screening of database for specific properties (De, Musil, Ingram, Baldauf, & Ceriotti, 2017; Hautier, 2019; Maier, Stöwe, & Sieg, 2007). Unfortunately, the dimensionality of these features (as well as most of other descriptors used in chemical and materials informatics) is very high, which makes the resulting classifications, clustering or mapping very hard to visualize. Dimensionality reduction algorithms (Ceriotti, Tribello, & Parrinello, 2011; McInnes, Healy, & Melville, 2018; Schölkopf, Smola, & Müller, 1998) can reduce the number of relevant dimensions to a handful, creating 2D or 3D maps of the full database.

The Journal of Open Source Software

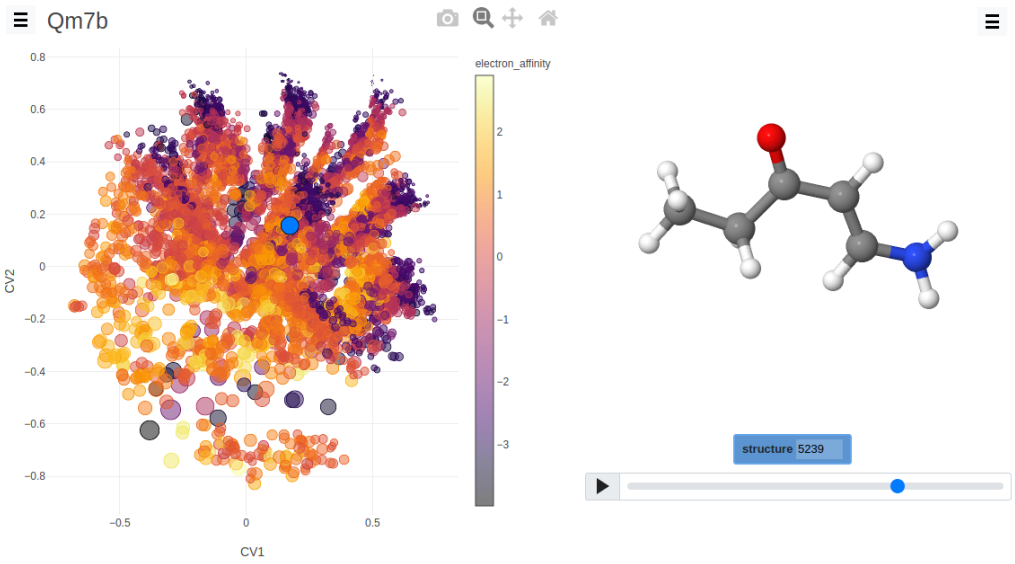

Figure 1: The Qm7b database (Montavon et al., 2013) visualized with chemiscope

Chemiscope is a graphical tool for the interactive exploration of materials and molecular databases, correlating local and global structural descriptors with the physical properties of the different systems. The interface consists of two panels. The left panel displays a 2D or 3D scatter plot, in which each point corresponds to a chemical entity. The axes, color, and style of each point can be set to represent a property or a structural descriptor to visualize structure-property relations directly. Structural descriptors are not computed directly by chemiscope, but must be obtained from one of the many codes implementing general-purpose atomic representation (librascal, n.d.; QUIP, n.d.) or more specialized descriptors. Since the most common descriptors can be very high dimensional, it can be convenient to apply a dimensionality reduction algorithm that maps them to a lower-dimensional space for easier visualization. For example the sketch-map algorithm (Ceriotti et al., 2011) was used with the Smooth Overlap of Atomic Positions representation (Bartók et al., 2013) to generate the visualization in Figure 1. The right panel displays the three-dimensional structure of the chemical entities, possibly including periodic repetition for crystals. Visualizing the chemical structure can help in finding an intuitive rationalization of the layout of the dataset and the structure-property relations.

Whereas similar tools (De & Ceriotti, 2019; Gong et al., 2013; Gütlein, Karwath, & Kramer, 2014; Probst & Reymond, 2017) only allow visualizing maps and structures in which each data point corresponds to a molecule, or a crystal structure, a distinctive feature of chemiscope is the possibility of visualizing maps in which points correspond to atom-centred environments. This is useful, for instance, to rationalize the relationship between structure and atomic properties such as nuclear chemical shieldings (Figure 2). This is also useful as a diagnostic tool for the many machine-learning schemes that decompose properties into atom-centred contributions (Bartók, Payne, Kondor, & Csányi, 2010; Behler & Parrinello, 2007).

The Journal of Open Source Software

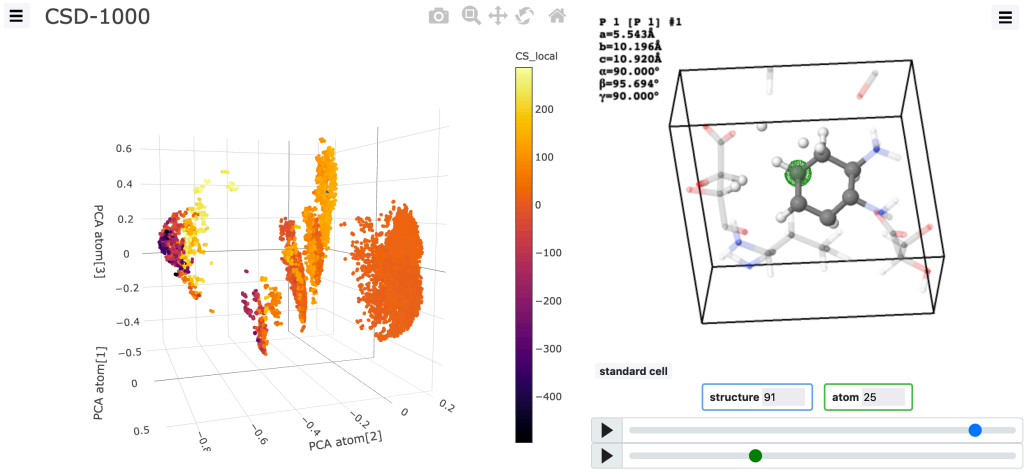

Figure 2: Database of chemical shieldings (Paruzzo et al., 2018) in chemiscope demonstrating the use of a 3D plot and highlighting of atomic environments

Chemiscope took strong inspiration from a previous similar graphical software, the interactive sketch-map visualizer (De & Ceriotti, 2019). This previous software was used in multiple research publication, related to the exploration of large-scale databases, and the mapping of structure-property relationships (De, Bartók, Csányi, & Ceriotti, 2016; De et al., 2017; Musil et al., 2018).

Implementation

Chemiscope is implemented using the web platform: HTML5, CSS and WebGL to display graphical elements, and TypeScript (compiled to JavaScript) for interactivity. It uses Plotly.js to render and animate 2D and 3D plots; and the JavaScript version of Jmol to display atomic structures. The visualization is fast enough to be used with datasets containing up to a million points, reacting to user input within a few hundred milliseconds in the default 2D mode. More elaborate visualizations are slower, while still handling 100k points easily.

The use of web technologies makes chemiscope usable from different operating systems without the need to develop, maintain and package the code for each operating system. It also means that we can provide an online service at http://chemiscope.org that allows users to visualize their own dataset without any local installation. Chemiscope is implemented as a library of re-usable components linked together via callbacks. This makes it easy to modify the default interface to generate more elaborate visualizations, for example, displaying multiple maps generated with different parameters of a dimensionality reduction algorithm. Chemiscope can also be distributed in a standalone mode, where the code and a predefined dataset are merged together as a single HTML file. This standalone mode is useful for archival purposes, for example as supplementary information for a published article and for use in corporate environments with sensitive datasets.

Acknowledgements

The development of chemiscope have been funded by the NCCR MARVEL, the MAX European centre of excellence, and the European Research Council (Horizon 2020 grant agreement no. 677013-HBMAP).

References

Bartók, A. P., Kondor, R., & Csányi, G. (2013). On representing chemical environments. *Physical Review B*, *87*(18). doi:10.1103/physrevb.87.184115

Bartók, A. P., Payne, M. C., Kondor, R., & Csányi, G. (2010). Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Physical Review Letters*, *104*(13). doi:10.1103/physrevlett.104.136403

Behler, J., & Parrinello, M. (2007). Generalized neural-network representation of high-dimensional potential-energy surfaces. *Physical Review Letters*, *98*(14). doi:10.1103/physrevlett.98.146401

Ceriotti, M., Tribello, G. A., & Parrinello, M. (2011). Simplifying the representation of complex free-energy landscapes using sketch-map. *Proceedings of the National Academy of Sciences*, *108*(32), 13023–13028. doi:10.1073/pnas.1108486108

De, S., Bartók, A. P., Csányi, G., & Ceriotti, M. (2016). Comparing molecules and solids across structural and alchemical space. *Physical Chemistry Chemical Physics*, *18*(20), 13754–13769. doi:10.1039/c6cp00415f

De, S., & Ceriotti, M. (2019). *Interactive sketchmap visualizer*. Zenodo. doi:10.5281/zenodo.3541831

De, S., Musil, F., Ingram, T., Baldauf, C., & Ceriotti, M. (2017). Mapping and classifying molecules from a high-throughput structural database. *Journal of Cheminformatics*, *9*(1). doi:10.1186/s13321-017-0192-4

Gong, J., Cai, C., Liu, X., Ku, X., Jiang, H., Gao, D., & Li, H. (2013). ChemMapper: A versatile web server for exploring pharmacology and chemical structure association based on molecular 3D similarity method. *Bioinformatics*, *29*(14), 1827–1829. doi:10.1093/bioinformatics/btt270

Gütlein, M., Karwath, A., & Kramer, S. (2014). CheS-mapper 2.0 for visual validation of (q)SAR models. *Journal of Cheminformatics*, *6*(1). doi:10.1186/s13321-014-0041-7

Hautier, G. (2019). Finding the needle in the haystack: Materials discovery and design through computational ab initio high-throughput screening. *Computational Materials Science*, *163*, 108–116. doi:10.1016/j.commatsci.2019.02.040

librascal. (n.d.). Retrieved from https://github.com/cosmo-epfl/librascal

Maier, W., Stöwe, K., & Sieg, S. (2007). Combinatorial and high-throughput materials science. *Angewandte Chemie International Edition*, *46*(32), 6016–6067. doi:10.1002/anie.200603675

McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform manifold approximation and projection for dimension reduction.

Montavon, G., Rupp, M., Gobre, V., Vazquez-Mayagoitia, A., Hansen, K., Tkatchenko, A., Müller, K.-R., et al. (2013). Machine learning of molecular electronic properties in chemical compound space. *New Journal of Physics*, *15*(9), 095003. doi:10.1088/1367-2630/15/9/095003

Musil, F., De, S., Yang, J., Campbell, J. E., Day, G. M., & Ceriotti, M. (2018). Machine learning for the structure-energy-property landscapes of molecular crystals. *Chemical Science*, *9*(5), 1289–1300. doi:10.1039/c7sc04665k

Paruzzo, F. M., Hofstetter, A., Musil, F., De, S., Ceriotti, M., & Emsley, L. (2018). Chemical shifts in molecular solids by machine learning. *Nature Communications*, *9*(1). doi:10.1038/s41467-018-06972-x

Probst, D., & Reymond, J.-L. (2017). FUn: A framework for interactive visualizations of large, high-dimensional datasets on the web. (J. Wren, Ed.)*Bioinformatics*, *34*(8), 1433–1435. doi:10.1093/bioinformatics/btx760

QUIP. (n.d.). Retrieved from http://libatoms.github.io/QUIP/

Ramakrishnan, R., Dral, P. O., Rupp, M., & Lilienfeld, O. A. von. (2014). Quantum chemistry structures and properties of 134 kilo molecules. *Scientific Data*, *1*(1). doi:10.1038/sdata.2014.22

Ruddigkeit, L., Deursen, R. van, Blum, L. C., & Reymond, J.-L. (2012). Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. *Journal of Chemical Information and Modeling*, *52*(11), 2864–2875. doi:10.1021/ci300415d

Schölkopf, B., Smola, A., & Müller, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, *10*(5), 1299–1319. doi:10.1162/089976698300017467

Willatt, M. J., Musil, F., & Ceriotti, M. (2019). Atom-density representations for machine learning. *The Journal of Chemical Physics*, *150*(15), 154110. doi:10.1063/1.5090481