
chemreps Documentation

Release 0.1.0

See AUTHORS.md

Dec 08, 2020

CONTENTS:

| | | |
|----------|--------------------------------|-----------|
| 1 | Current Implementations | 3 |
| 2 | Representation requests | 5 |
| 3 | Disclaimers | 7 |
| 4 | For help | 9 |
| 5 | Citing | 11 |
| 6 | Indices and tables | 15 |

chemreps is a Python package for the creation of molecular representations for the purpose of machine learning. The molecular representations included in this library are implemented/adapted from current literature. The aim of chemreps is to provide an easy to use library for making molecular representations that can be then used with machine learning packages such as Scikit-Learn and Tensorflow.

CURRENT IMPLEMENTATIONS

- Coulomb Matrix
- Bag of Bonds
- Bonds/Nonbonding, Angles, Torsions
- Just Bonds
- Morgan Fingerprints (RDKit Dependency)

The citations for the literature from which the representations are implemented/adapted from can be found in the source code for each representation.

REPRESENTATION REQUESTS

Requests for new representations to be added can be made by raising an issue and labeling it as a feature request. Before requesting a new representation, please check under the Representation project in the Projects tab to see if that representation is included in the current work or progress.

DISCLAIMERS

- These are attempts at the recreation of molecular representations from literature and may not be implemented properly.
- If we do not implement something properly, please make an issue to let us know.
- This is solely a representation library and will not perform machine learning.

FOR HELP

If you need any help using chemreps, feel free to post in our [Gitter](#).

We now have a [Zenodo release](#)!

5.1 Install

5.1.1 Installation

The latest release version can be installed with:

```
pip install chemreps
```

The latest development version can be installed by:

```
git clone https://github.com/chemreps/chemreps
cd chemreps
pip install -e .
```

5.1.2 Dependencies

chemreps requires:

- Python (≥ 3.6)
- NumPy (≥ 1.12)
- cclib (≥ 1.5)
- QCElemental

5.1.3 Optional Dependencies

- RDKit

5.1.4 Testing

Tests can be run in the top-level directory with the command `pytest -v --cov=chemreps tests/`

5.2 Examples

Examples for how to use chemreps can also be found on GitHub under [chemreps/examples](#) as well as tried out on [Binder](#).

5.2.1 Coulomb Matrix

The Coulomb Matrix representation was designed for the modeling of atomization energies and as a result requires information needed by the Hamiltonian, namely the set of Cartesian coordinates R_I and nuclear charges Z_I . With this information the Coulomb matrix M is built

where the off-diagonal elements correspond to the Coulomb repulsion between I and J . The expression for the diagonal elements is the result of an encoded polynomial fit to the atomic energies and nuclear charge.

- DOI: 10.1103/PhysRevLett.108.058301

A Coulomb Matrix representation can be made below by importing `coulomb_matrix` from `chemreps.coulomb_matrix` and passing molecule file and *size* of the matrix as seen below:

```
from chemreps.coulomb_matrix import coulomb_matrix

mfile = '../data/sdf/butane.sdf'
cm = coulomb_matrix(mfile, size=14)
```

5.2.2 Bag of Bonds & other bagged representations

Bagged representations require the size of the bags and empty bags to be passed to the representation. The *bag_sizes* and *bags* can be made using *BagMaker* from `chemreps.bagger` if they are unknown or not yet made.

```
from chemreps.bagger import BagMaker

dataset = '../data/sdf/'
bagger = BagMaker('BoB', dataset)
bags = bagger.bags
bag_sizes = bagger.bag_sizes
```

Once the bags have been made, they can be passed to the representation function. Bag of Bonds (BoB) will be shown for this example, so *bag_of_bonds* is passed the newly created bags to create a BoB representation for the *butane.sdf*.

```
from chemreps.bag_of_bonds import bag_of_bonds

mfiles = '../data/sdf/butane.sdf'
rep = bag_of_bonds(mfiles, bagger.bags, bagger.bag_sizes)
```


5.2.3 Further Questions

Further questions on how to use *chemreps* to make molecular representations can be asked on [Gitter](#).

5.3 Contributing

Work in progress.

5.4 chemreps package

Contents

- *chemreps* package
 - *Submodules*
 - *chemreps.bagger* module
 - *chemreps.bag_of_bonds* module
 - *chemreps.bat* module
 - *chemreps.coulomb_matrix* module
 - *chemreps.dataset* module
 - *chemreps.fingerprints* module
 - *chemreps.just_bonds* module

5.4.1 Submodules

chemreps.utils package

Contents

- *chemreps.utils* package
 - *chemreps.utils.bag_handler* module
 - *chemreps.utils.calcs* module
 - *chemreps.utils.molecule* module

chemreps.utils.bag_handler module

chemreps.utils.calcs module

chemreps.utils.molecule module

5.4.2 chemreps.bagger module

5.4.3 chemreps.bag_of_bonds module

5.4.4 chemreps.bat module

5.4.5 chemreps.coulomb_matrix module

5.4.6 chemreps.dataset module

5.4.7 chemreps.fingerprints module

5.4.8 chemreps.just_bonds module

INDICES AND TABLES

- `genindex`
- `modindex`
- `search`