# TS-DiffuGen

*Release 2023*

**Sacha Raffaud**

**Aug 31, 2023**

# CONTENTS

CHAPTER

ONE

# EXAMPLE FUNCTIONS AND CLASSES FROM THE PACKAGE INCLUDE:

See *src.evaluate_samples* for details.

src.evaluate_samples.**calc_cov_mat**(*results_matrix*, *cov_threshold=0.1*)

Calculate COV and MAT scores based on D-MAE matrix.

**Parameters**

- **results_matrix** (*np.ndarray*) – D-MAE matrix.
- **cov_threshold** (*float, optional*) – COV threshold. Defaults to 0.1.

**Returns**

Calculated MAT-R mean, median, and COV-R scores.

**Return type**

tuple

src.evaluate_samples.**calculate_DMAE**(*gen_mol*, *true_mol*)

Calculate D-MAE between inter-atomic distance matrices.

**Parameters**

- **gen_mol** (*list*) – Inter-atomic distance matrix of generated molecule.
- **true_mol** (*list*) – Inter-atomic distance matrix of true molecule.

**Returns**

D-MAE value.

**Return type**

float

src.evaluate_samples.**calculate_best_rmse**(*gen_mol*, *ref_mol*, *max_iters=100000*, *use_hydrogens=False*)

Calculate Best RMSD between RDKit Molecule Objects.

**Parameters**

- **gen_mol** (*Chem.Mol*) – RDKit molecule object representing generated molecule.
- **ref_mol** (*Chem.Mol*) – RDKit molecule object representing reference molecule.
- **max_iters** (*int, optional*) – Maximum atom matches. Defaults to 100_000.
- **use_hydrogens** (*bool, optional*) – True to include hydrogens. Defaults to False.

**Returns**

Best RMSD value.

**Return type**
float

src.evaluate_samples.**calculate_distance_matrix**(*coordinates*)

Calculate pairwise distance matrix from 3D coordinates.

**Parameters**
**coordinates** (*list*) – List of 3D coordinates for each atom.

**Returns**
Pairwise distance matrix.

**Return type**
np.ndarray

src.evaluate_samples.**create_lists**(*original_path*, *RMSD=False*)

Create lists of true and generated molecules from the given path.

**Parameters**

- **original_path** (*str*) – Path to the original directory containing molecule files.
- **RMSD** (*bool, optional*) – True if RMSD format, False if standard XYZ format.

**Returns**
Two lists containing RDKit molecule objects.

**Return type**
tuple

src.evaluate_samples.**create_table**(*true_mols*, *gen_mols*, *max_iters=1*, *metric='RMSD'*)

Create comparison table between molecules.

**Parameters**

- **true_mols** (*list*) – List of true molecule RDKit objects.
- **gen_mols** (*list*) – List of lists containing generated molecule RDKit objects.
- **max_iters** (*int*) – Maximum atom matches for RMSD calculation.
- **metric** (*str*) – Metric choice, "RMSD" or "DMAE".

**Returns**
DataFrame of comparison metrics.

**Return type**
pd.DataFrame

src.evaluate_samples.**evaluate**(*sample_path*, *evaluation_type*, *cov_threshold=0.1*)

Evaluate generated samples using RMSE or D-MAE metrics.

**Parameters**

- **sample_path** (*str*) – Path to sample directory.
- **evaluation_type** (*str*) – Metric choice, "RMSD" or "DMAE".
- **cov_threshold** (*float, optional*) – COV threshold. Defaults to 0.1.

src.evaluate_samples.**get_paths**(*sample_path*)

Load molecule files from a sample path and organize them into true and generated samples.

**Parameters**
**sample_path** (*str*) – Path to the sample directory containing molecule files.

**Returns**
Two lists containing true and generated sample file paths.

**Return type**
tuple

src.evaluate_samples.**import_xyz_file**(*molecule_path*, *RMSD=False*)

Import an XYZ file as an RDKit molecule object.

**Parameters**

- **molecule_path** (*str*) – File path of the XYZ file to be imported.
- **RMSD** (*bool, optional*) – True if RMSD format, False if standard XYZ format.

**Returns**
RDKit molecule object or None if loading fails.

**Return type**
Chem.Mol or None

# PYTHON MODULE INDEX

## S

# INDEX