
Kyos Documentation

Release 0.2.0

Nathan Xue

May 13, 2020

Contents

1	Kyos	3
1.1	Features	3
1.2	Citing Kyos	3
1.3	License	3
2	Installation	5
2.1	Upgrading Kyos	5
2.2	Uninstalling Kyos	5
3	Usage	7
4	Contributing	9
4.1	Types of Contributions	9
4.2	Get Started!	10
4.3	Pull Request Guidelines	11
4.4	Tips	11
5	Credits	13
5.1	Development Lead	13
5.2	CFSAN Bioinformatics Team	13
5.3	External Contributors	13
6	History	15
6.1	0.2.0 (2020-05-13)	15
6.2	0.1.0 (2018-12-07)	15
7	Indices and tables	17

Contents:

Tools for haploid variant calling with Deep Neural Networks.

Kyos was developed by the United States Food and Drug Administration, Center for Food Safety and Applied Nutrition.

- Free software
- Documentation: <https://kyos.readthedocs.io>
- Source Code: <https://github.com/CFSAN-Biostatistics/kyos>
- PyPI Distribution: <https://pypi.python.org/pypi/kyos>

1.1 Features

- Extract features from a BAM file
- Train and Test a neural network

1.2 Citing Kyos

To cite Kyos, please reference the Kyos GitHub repository:

<https://github.com/CFSAN-Biostatistics/kyos>

1.3 License

See the LICENSE file included in the Kyos distribution.

CHAPTER 2

Installation

At the command line:

```
$ pip install --user kyos
```

Update your `.bashrc` file with the path to user-installed python packages:

```
export PATH=~/.local/bin:$PATH
```

Or, if you have `virtualenvwrapper` installed:

```
$ mkvirtualenv kyos  
$ pip install kyos
```

2.1 Upgrading Kyos

If you previously installed with `pip`, you can upgrade to the newest version from the command line:

```
$ pip install --user --upgrade kyos
```

2.2 Uninstalling Kyos

If you installed with `pip`, you can uninstall from the command line:

```
$ pip uninstall kyos
```


CHAPTER 3

Usage

Before you can use Kyos to call variants, you will need to prepare the input datasets. The workflow begins with one or more BAM files. If you don't already have BAM files, you could use the [CFSAN SNP Pipeline](#) to create the BAM files.

When extracting features from BAM files, you will need to supply the known-truth if you intend to use the tabulated features for training and testing the neural network. The `-t` command line option to the `tabulate` command adds an extra `Truth` column to the output tsv file.

Kyos is currently dependent upon [SNP Mutator](#) to generate the known-truth datasets for supervised learning. A future version will use VCF files instead.

To extract tabular data from a BAM file:

```
kyos tabulate -t TRUTH_FILE input.bam output.tsv ref.fasta
```

To merge multiple tabulated files:

```
kyos merge file1.tsv file2.tsv file3.tsv ... > train.tsv
```

To train a neural network model:

```
kyos train train.tsv validate.tsv model.h5
```

To test a neural network model:

```
kyos test model.h5 test.tsv
```


Contributions are welcome, and they are greatly appreciated! Every little bit helps, and credit will always be given. You can contribute in many ways:

4.1 Types of Contributions

4.1.1 Report Bugs

Report bugs at <https://github.com/CFSAN-Biostatistics/kyos/issues>.

If you are reporting a bug, please include:

- Your operating system name and version.
- Any details about your local setup that might be helpful in troubleshooting.
- Detailed steps to reproduce the bug.

4.1.2 Fix Bugs

Look through the GitHub issues for bugs. Anything tagged with “bug” is open to whoever wants to implement it.

4.1.3 Implement Features

Look through the GitHub issues for features. Anything tagged with “feature” is open to whoever wants to implement it.

4.1.4 Write Documentation

Kyos could always use more documentation, whether as part of the official Kyos docs, in docstrings, or even on the web in blog posts, articles, and such.

4.1.5 Submit Feedback

The best way to send feedback is to file an issue at <https://github.com/CFSAN-Biostatistics/kyos/issues>.

If you are proposing a feature:

- Explain in detail how it would work.
- Keep the scope as narrow as possible, to make it easier to implement.
- Remember that this is a volunteer-driven project, and that contributions are welcome :)

4.2 Get Started!

Ready to contribute? Here's how to set up *kyos* for local development.

1. Fork the *kyos* repo on GitHub.
2. Clone your fork locally:

```
$ git clone git@github.com:your_name_here/kyos.git
```

3. Install your local copy into a virtualenv. Assuming you have virtualenvwrapper installed, this is how you set up your fork for local development:

```
$ mkvirtualenv kyos
$ cd kyos/
$ pip install sphinx_rtd_theme      # the documentation uses the ReadTheDocs theme
$ pip install pytest
$ python setup.py develop
```

4. Create a branch for local development:

```
$ git checkout -b name-of-your-bugfix-or-feature
```

Now you can make your changes locally.

5. When you're done making changes, check that your changes pass flake8 and the tests, including testing other Python versions with tox:

```
$ flake8 kyos tests
$ pytest -v
$ tox
```

To get flake8 and tox, just pip install them into your virtualenv.

6. Update the documentation and review the changes locally with sphinx:

```
$ cd docs
$ sphinx-build -b html . ../_build
$ xdg-open _build/index.html
```

7. Commit your changes and push your branch to GitHub:

```
$ git add .  
$ git commit -m "Your detailed description of your changes."  
$ git push origin name-of-your-bugfix-or-feature
```

8. Submit a pull request through the GitHub website.

4.3 Pull Request Guidelines

Before you submit a pull request, check that it meets these guidelines:

1. The pull request should include tests.
2. If the pull request adds functionality, the docs should be updated. Put your new functionality into a function with a docstring, and add the feature to the list in README.rst.
3. The pull request should work for Python 2.7, 3.4, 3.5, 3.6, and for PyPy.

4.4 Tips

To run a subset of tests:

```
$ pytest -v tests/test_kyos.py
```


5.1 Development Lead

- Nathan Xue <xue.nathanv1.0@gmail.com>

5.2 CFSAN Bioinformatics Team

- Nathan Xue <xue.nathanv1.0@gmail.com>

5.3 External Contributors

None yet. Why not be the first?

6.1 0.2.0 (2020-05-13)

- Changes for Tensorflow 2.

6.2 0.1.0 (2018-12-07)

- Initial version.

CHAPTER 7

Indices and tables

- `genindex`
- `search`