joblib 0.10.3.dev0

Lightweight pipelining: using Python functions as pipeline jobs. **Downloads ↓**

Package Documentation

Joblib is a set of tools to provide **lightweight pipelining in Python**. In particular, joblib offers:

1. transparent disk-caching of the output values and lazy re-evaluation (memoize pattern)
2. easy simple parallel computing
3. logging and tracing of the execution

Joblib is optimized to be **fast** and **robust** in particular on large data and has specific optimizations for *numpy* arrays. It is **BSD-licensed**.

User documentation: **http://pythonhosted.org/joblib**
Download packages: **http://pypi.python.org/pypi/joblib#downloads**
Source code: **http://github.com/joblib/joblib**
Report issues: **http://github.com/joblib/joblib/issues**

Vision

The vision is to provide tools to easily achieve better performance and reproducibility when working with long running jobs.

- **Avoid computing twice the same thing**: code is rerun over an over, for instance when prototyping computational-heavy jobs (as in scientific development), but hand-crafted solution to alleviate this issue is error-prone and often leads to unreproducible results
- **Persist to disk transparently**: persisting in an efficient way arbitrary objects containing large data is hard. Using joblib's caching mechanism avoids hand-written persistence and implicitly links the file on disk to the execution context of the original Python object. As a result, joblib's persistence is good for resuming an application status or computational job, eg after a crash.

Joblib strives to address these problems while **leaving your code and your flow control as unmodified as possible** (no framework, no new paradigms).

Main features

1. **Transparent and fast disk-caching of output value:** a memoize or make-like functionality for Python functions that works well for arbitrary Python objects, including very large numpy arrays. Separate persistence and flow-execution logic from domain logic or algorithmic code by writing the operations as a set of steps with well-defined inputs and outputs: Python functions. Joblib can save their computation to disk and rerun it only if necessary:

```
>>> from joblib import Memory
>>> mem = Memory(cachedir='/tmp/joblib')
>>> import numpy as np
>>> a = np.vander(np.arange(3)).astype(np.float)
>>> square = mem.cache(np.square)
>>> b = square(a)                                   # doctest: +ELLIPSIS
________________________________________________________________________________
[Memory] Calling square...
square(array([[ 0.,  0.,  1.],
       [ 1.,  1.,  1.],
       [ 4.,  2.,  1.]]))
_____________________________________________________________square - 0...s, 0.0min

>>> c = square(a)
>>> # The above call did not trigger an evaluation
```

2. **Embarrassingly parallel helper:** to make it easy to write readable parallel code and debug it quickly:

```
>>> from joblib import Parallel, delayed
>>> from math import sqrt
>>> Parallel(n_jobs=1)(delayed(sqrt)(i**2) for i in range(10))
[0.0, 1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0]
```

3. **Logging/tracing:** The different functionalities will progressively acquire better logging mechanism to help track what has been ran, and capture I/O easily. In addition, Joblib will provide a few I/O primitives, to easily define logging and display streams, and provide a way of compiling a report. We want to be able to quickly inspect what has been run.

4. **Fast compressed Persistence**: a replacement for pickle to work efficiently on Python objects containing large data (*joblib.dump* & *joblib.load*).

File	Type	Py Version	Uploaded on	Size
joblib-0.10.3.dev0-py2.py3-none-any.whl (**md5**)	Python Wheel	3.5	2016-09-06	162KB
joblib-0.10.3.dev0-py3.5.egg (**md5**)	Python Egg	3.5	2016-09-06	291KB
joblib-0.10.3.dev0.tar.gz (**md5**)	Source		2016-09-06	485KB

Author: Gael Varoquaux
Documentation: joblib package documentation
Home Page: http://pythonhosted.org/joblib/
License: BSD
Platform: any
Categories
- **Development Status :: 5 - Production/Stable**
- **Environment :: Console**
- **Intended Audience :: Developers**
- **Intended Audience :: Education**
- **Intended Audience :: Science/Research**
- **License :: OSI Approved :: BSD License**
- **Operating System :: OS Independent**
- **Programming Language :: Python :: 2.6**
- **Programming Language :: Python :: 2.7**
- **Programming Language :: Python :: 3**
- **Programming Language :: Python :: 3.3**
- **Programming Language :: Python :: 3.4**
- **Topic :: Scientific/Engineering**
- **Topic :: Software Development :: Libraries**
- **Topic :: Utilities**

Package Index Owner: GaelVaroquaux, ogrisel, lesteve
Package Index Maintainer: lesteve
DOAP record: joblib-0.10.3.dev0.xml