ADAMS

Advanced **D**ata mining **A**nd **M**achine learning **S**ystem

Module: adams-tensorflow



Peter Reutemann

January 8, 2020

©2018-2019



THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato



Except where otherwise noted, this work is licensed under
http://creativecommons.org/licenses/by-sa/4.0/

Contents

1 TensorFlow **5**

2 Flow **7**

Bibliography **9**

Chapter 1

TensorFlow

TensorFlow[2] is an open source software library for high performance numerical computation. Its flexible architecture allows easy deployment of computation across a variety of platforms (CPUs, GPUs, TPUs), and from desktops to clusters of servers to mobile and edge devices. Originally developed by researchers and engineers from the Google Brain team within Google's AI organization, it comes with strong support for machine learning and deep learning and the flexible numerical computation core is used across many other scientific domains.

Chapter 2

Flow

In order to use Tensorflow's Object Detection API[3], it is necessary to create tfrecords[4] from your annotated data. Generating the tfrecords can be achieved with the *wai.tfrecords*[5] library. But rather than having to split the data manually in train and test sets, you can use ADAMS' functionality to separate file lists into subsests (e.g., using *PrepareFileBasedDataset*).

To set up the wai.tfrecords library, create a virtual environment (Python 3.7+) and install the requirements from *adams-tensorflow-generate_tfrecords-requirements.txt*.
For instance, on Linux run these commands (adjust the path accordingly):

```
virtualenv -p /usr/bin/python3.7 venv
./venv/bin/pip install -r adams-tensorflow-generate_tfrecords-requirements.txt
```

With the virtual environment in place, you can run the following flows:

- *adams-tensorflow-generate_tfrecords-cv.flow* – generates cross-validation folds
- *adams-tensorflow-generate_tfrecords-split.flow* – generates train/test split

When prompted by the flows to supply a Python executable, simply select the executable from within the previously generated virtual environment. The flow will construct file lists from the input data and call the conversion script to generate tfrecord files (and label/file lists) from these.

Bibliography

[1] *ADAMS* – Advanced Data mining and Machine learning System
`https://adams.cms.waikato.ac.nz/`

[2] *TensorFlow* – an open source machine learning framework for everyone.
`https://www.tensorflow.org/`

[3] *Object Detection API* – The TensorFlow Object Detection API is an open source framework built on top of TensorFlow that makes it easy to construct, train and deploy object detection models.
`https://github.com/tensorflow/models/tree/master/research/object_detection`

[4] *tfrecord* – The TFRecord format is a simple format for storing a sequence of binary records.
`https://www.tensorflow.org/tutorials/load_data/tfrecord`

[5] *wai.tfrecords* – A Python library to generate TFRecord files (based on Object Detection API code).
`https://github.com/waikato-datamining/tensorflow/tree/master/tfrecords`