M05 - Mini project

Human Activity Recognition from Continuous Ambient Sensor Data

Steve Devènes, Amara Spano

Unidistance

October 12, 2020

Working hypothesis

It's possible to perform human activity recognition from continuous ambient sensor data.

Data

- Data is available online on the UC Irvine machine learning repository(Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.).
- Data are downloaded through the command line.
- The dataset is large, it contains 36 features measured plus one output for the classification label of the activity (35 different activities), for a total of 13956534 entries (30 different participants).
- For the experiments, data from a unique house is used
- There are two evaluation protocols splitting the data into 80% train and 20% test sets with 2 different random seeds.

Workflow

For different experiments:

1. Load the training data
2. Create and train a random forest classifier
3. Load the test data
4. Make prediction on test data
5. Print the confusion matrix for the model evaluation. Confusion matrices are also available in graphs with plotly.express.

The experiments run sequentially.

Version control

On github at `https://github.com/sdevenes/M05_MiniProject`
The work is organized using github issues to create and assign tasks.
The general approach:

- Create 1 branch per feature named *feature/feature_name* or *issue_#issue/issue_name*
- When the feature is complete, do a pull request with the other as a reviewer

clarified docstring for database.py	21 sept. 2020 1:05	Amara Spano <amar	1066175
docstring for paper.py	21 sept. 2020 1:03	Amara Spano <amar	9bf55bf
docstring for database.py	21 sept. 2020 0:53	Amara Spano <amar	f08a0fa
implemented test tree depth and number of trees	21 sept. 2020 0:11	Amara Spano <amar	dd2a6b9
added papers.py and implemented basic test + co	20 sept. 2020 23:54	Amara Spano <amar	05406cd
fixed deprecated warning in database.py	20 sept. 2020 23:12	Amara Spano <amar	b8aab31
Merge pull request #15 from sdevenes/feature/co	20 sept. 2020 22:09	spanoamara <32494	d6480b7
Merge pull request #13 from sdevenes/feature/alg	20 sept. 2020 22:09	spanoamara <32494	1f3742b
add automatic code documentation using sphinx	20 sept. 2020 11:25	SteveDevenes <steve	002f5c2
replace confusion matrix plot using plotly.express	14 sept. 2020 11:47	SteveDevenes <steve	2bdaf16
add analysis and test script	8 sept. 2020 16:43	SteveDevenes <steve	66cecef
add random forest algo script	8 sept. 2020 16:42	SteveDevenes <steve	272f71f
Merge pull request #2 from sdevenes/feature/split	8 sept. 2020 14:08	Steve Devènes <stev	a018adc

Unit Testing and CI

- The majority of the functions in the code are covered by unit-tests for a total coverage of 86%, this was done using Nose python package. `https://coveralls.io/github/sdevenes/M05_MiniProject?branch=master`
- These tests are ran at each commit through travis CI: `https://travis-ci.org/github/sdevenes/M05_MiniProject` Procuring a quick detection of bugs and e_ors that could appears during the project development.
- Black is used to enforce formatting that conforms to PEP 8. The formatting is also tested with travis CI.

Documentation

Each function is commented with a docstring. The documentation is then build automatically in the CI using Sphinx at each new commit on the master branch.
Doc available here:
`https://sdevenes.github.io/M05_MiniProject/index.html`

Packaging and Deployment

- The package is deployed under *rr_sdas* version 0.2.0 on pypitest. `https://test.pypi.org/project/rr-sdas/0.2.0/`
- The package is composed of 3 sub-packages:
  1. download_data
  2. experiment
  3. tests
- The experiments are configurable through:
  1. Command line arguments (file sources and destination)
  2. *experiment.ini* for the random forest parameters
- The project is licensed under an MIT license as it is an open source license and it is less restrictive than a GPL license.