Code Guide Readme

Ninghui Li, Anqi Chen, Zitao Li, Tianhao Wang
Team DPSyn, Purdue University, Peking University
{ninghui, zitaoli, tianhaowang}@purdue.edu, caq@pku.edu.cn

Abstract

This file gives a statement of the mappings between steps in algorithm writeup and the analogous steps in the source code repository.

experiment.py

- the main program for users to run with input files and other parameters

data/DataLoader.py

- load config file which determines the identifier and binning settings
- load parameter file which describes the schema of original dataset by denoting the data type and valid values
- bin some numerical attributes
- encode the attributes to facilitate processing

method/dpsyn.py

- it inherits the Class synthesizer and utilize the library code in lib_dpsyn to synthesize records

method/synthesizer.py

- it mainly contains the functions to generate noisy marginals

lib_dpsyn/record_synthesizer.py

- the lib code for synthesizing records

lib_dpsyn/consistent.py

- the lib code for getting consistent marginals, i.e, make sure the noisy marginals submit to the dependency relationship, with the same summed count and non-negative counts

lib_dpsyn/view.py

- serves for data structure to deal with marginals

utils/advanced_composition.py

- include some functions for calculating noise type and corresponding parameters