

CatchCore: Catching Hierarchical Dense Subtensor

April-2019

Wenjie Feng (fengwenjie@ict.ac.cn / wenchiefeng.us@gmail.com)

1. General Information

- Python
- Version: 1.0
- Date: April-2019
- Author: Wenjie Feng (wenchiefeng.us@gmail.com)

2. Introduction

CatchCore is a novel framework to detect hierarchical dense cores in multi-aspect data (i.e. tensors). **CatchCore** has following properties:

- **unified metric**: provides a gradient-based optimized framework to detect dense blocks with theoretical guarantees as well as the MDL (Minimum Length Description) measurement
- **accuracy**: provides high accuracy in both synthetic and real datasets
- **effectiveness**: spots anomaly patterns and hierarchical high-dense community structure
- **scalability**: scales almost linearly with all input factors of tensor, also has linearly space complexity.

Detailed information about the methods is explained in the following paper.

- Wenjie Feng, Shenghua Liu, Huawei Shen, and Xueqi Cheng. “*CatchCore: Catching Hierarchical Dense Subtensor*”, The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD), 2019.

3. Installation

- This package requires python 3.6, to install some dependency library packages please type `pip install -r requirements`
- For demo (optional) for scratch graph features, type `make`.
- For packaging (optional), type `make tar`.

4. Input File Format

The input file lists all the tuples in a relation. Each line corresponding to a tuple and consists of the dimension attributes values, which are separated by a comma. For the measure attribute value, which can be given in the file with integer formant in the last columns of each rows, if it is not provided, the default value is '1'.

example.tensor is an example of the input file.

5. Output File Format

The output file lists the result of hierarchical dense subtensors.

- The '%' commented lines give the information about:
1st line --- the number of hierarchies ***K*** detected from the data and the dimension of tensor ***N***;
2nd line --- the dimension of each resultant hierarchies dense subtensors.
- The following ***K * N*** lines list the attribute values of each dimension of each hierarchical subtensor with format that: $k_n: a_1, a_2, \dots, a_t$ denoting that a_1, a_2, \dots, a_t are attribute values of n-th dimension of the k-th subtensor.

output/hierway.out contains the examples of the output file. Other running information, including density, are printed in the console.

The MDL measure result will be printed in the console rather than file.

6. Running

For any *run_*.py* script,

See help for usage and parameter explanation by typing:

python run_.py -h* or *python run_*.py --help*

See **Example** usage in *demo.sh* or the ***doc string*** in each *run_*.py*.