

# The usage of the BIDE algorithm

Prepared by Jianyong Wang, Email: [jianyong@tsinghua.edu.cn](mailto:jianyong@tsinghua.edu.cn)

## 1. System usage

### 1.1 Input parameters

1st argument: The specification file of the dataset

2nd argument: Relative support in decimal

Usage example: `bide bide_gaz.spec 0.0002`

Where `bide` is the executable file name, `bide_gaz.spec` is the specification file of the sequence dataset being mined, 0.0002 is the relative support.

As to the dataset file format, see sections 1.3 and 1.4.

### 1.2 Output

The discovered frequent sequences are printed into a file called “frequent.dat”.

Each line in the result file, “frequent.dat”, contains a frequent sequence in the form:

*event<sub>1</sub> event<sub>2</sub> ... event<sub>n</sub> : absolute support*

Here is an example:

6 24 748 : 66

### 1.3 Specification file format

The first line is the dataset file name, the second line is the number of unique items, the third line is the number of sequences, the fourth line is the maximal length of a sequence, and the fifth line is the average length of a sequence.

### 1.4 Dataset file format

Usually a sequence database consists of a series of sequences (strictly speaking, here a sequence is a string in the current implementation). Each line represents a sequence and ends with -1, and the entire dataset ends with -2. Here is a sample sequence:

38 81 256 399 756 841 962 1009 -1

Example datasets:

Gazelle dataset: `bide_gaz.spec` and `bide_gaz.txt`

Mcseq dataset: `mcseq.spec` and `mcseq.txt`