

Code user manual for the research article – LBSA Driver

Content

1. Brief Description -----	1
2. Parameters Description -----	1
3. Functions Description -----	2

1. Brief Description

The LBSA program, is a data science project written fully in the python programming language. This code is implemented to derive cancer driver genes from large datasets of cancer genomic data. The code is implemented on 4 TCGA datasets: GBM, BRCA, OV, and PRAD, as well as on a biological interaction network dataset: Reactome_FIsInGene_2021. The project consists of 3 folders, /Data, /Functions, and /Output folders.

- /Data folder consists of datasets used in the project.
- /Functions folder consists of the code for working functions used in the project implementation
- /Output folder contains all results from the implementation.

2. Parameters Description

- *Input parameters*
 - -n, --network, required=True, type=str, help=mutations file: the network parameter is used to input the network used in the algorithm.
 - -d, --dataset, required=True, type=str, help=dataset name: the dataset parameter is used for taking the datasets as input to the algorithm.
 - -c, --cooling, default=(1 - (10 ** -2)), type=int, help=Cooling factor: The cooling factor parameter is also an input parameter which is used as one of the stopping criterion for the LBSA algorithm.

- -i, --iterations, default=(10 ** 5), type=int, help=Number of iterations: number of iteration is also an input parameter used as one of the stopping criterion for the LBSA algorithm.
- *Extra parameters*
 - Lbsa_starttime: algorithm start timer
 - Lbsa_stoptime: algorithm end time

The starttime and end time are used to calculate the total running time of the algorithm.

3. Functions Description

The functions in this implementation are separated into different files. The function files include: LBSA.py, construct_bipartite_graph.py, construct_mutation_matrix.py, and results.py.

- LBSA.py: This file contains various functions which are all involved in the LBSA implementation:
 - calc_coverage: This function is used to calculate the coverage of a give gene set.
 - Get_parameters: This function is used for assigned respective input parameters
 - Generate_Temperature_List: Used to Generate the temperature list. It returns a list of temperatures in the range (0, 1000]
 - List_Based_Simulated_Annaeling: This is the function which actually implements the LBSA algorithm. This function returns the list of drivers.
- construct_bipartite_graph.py: This file contains the function to generate the bipartite graph.
 - Create_bipartite_graph: this function takes 3 parameters as input: mutation_matrix, influence graph and gene_expression_matrix, and returns a bipartite graph.
- Construct_mutation_matrix.py: This file contains a function called the get_mutation_matrix_from_mat which is used to create a mutation matrix used as the input when creating the bipartite graph.

- Results.py: This file contains the report functions used for outputting and displaying the results of the algorithm.