This directory contains the Heterogeneous co-embedding code(LEDF) and the instruction for how to use it to obtain the each result in our paper. The parameter setting used in this file is exactly the same as that we used to produce the results.

To run the CODE&ACAS algorithms, please refer to the ACAS_release directory.

TM TM TM TM TM TM TM TM TM TM TM TM TM TM TM TM TM TM TM TM ***Arguments Description***

Input argument is a cell and its arguments are described as follow:

R —— m x n matrix, relation values between two sets of objects

dim —— integer, the embedded dimensionality

method —— String, the algorithm name, only included one option of the proposed algorithm: LEDeig

model_optimisation —— Cell, options for optimisation
Default: `{'optimisation', 'grid', 'obj_option', 'quant_knn', 'obj_para', {'knn', 10,10}}`;

optimisation —— String, chosen optimisation scheme.including ga and grid. Default 'ga'
obj_option —— String, choosing options for computing the optimisation score. Options:
obj_para —— Cell, `{'knn', kr,kc}`, where kr, kc are the number of neighbours used by
quant_knn score,these two parameters are used to compute the K(R) and K(Q)
in the paper.
parameter_range —— Cell, parameter searching range for the proposed algorithm. Including the eta_1, eta_2,
alpha (correspond to xi in the paper), beta (correspond to gamma in the paper). The
following parameter range is used for all data:
`{'eta1',[-1, 10],'eta2', [-1, 10],'alpha',[0, 3], 'beta',[0,3]}`

Output argument is a structure including the following field:

X —— m x k matrix, embeddings for the row objects in R
Y —— n x k matrix, embeddings for the column objects in R
model —- Structure, contains the 4 parameter values (eta1, eta2, alpha, beta) for the computed co-embedding

™ ***Example results***

Synthetical Data

At first load the associate synthetical data to the workspace from the Data folder, for example:

**********code**********

```
rng(213);
% load the associate data file
load('Data/dot.mat');
```

**********code**********

The **Plotcluster.m** function is used to plot data in 2&3 dimensions. To use this function, run the following code to see the plot:

**********code**********

```
figure
Plotcluster(X', lx, Y', ly);
```

**********code**********

Thus, we can able to visualising the X and Y data points with their coloured labeling information corresponding to label vector lx and ly, respectively. The synthetical relational matrix **R** is computed by the **CalRelationXY.m** function:

**********code**********

```
R    = CalRelationXY(X,Y,'gaussian','mean');
```

**********code**********

At last , we obtain the row and column co-embeddings of LEDF as well as plot it by running script of the following:

**********code**********

```
opt  = {'optimisation', 'ga', 'obj_option', 'quant_knn', 'obj_para',
{'knn', 10,10}};

para_range = {'eta1',[-1, 10],'eta2', [-1, 10],'alpha',[0, 3], 'beta',
[0,3]};

I    = {'R', R, 'method', 'LEDeig','dim',2, 'model_optimisation', opt,
'parameter_range',para_range };

led    = LEDF(I);
led    = training(led);


figure;
Plotcluster(led.X',lx,led.Y',ly);
```

**********code**********

Small Clinical Trial Text Data

This is a smaller database we extracted from the Clinical Trial data which only include 4 topics domain, and the resulting co-embedding we used in the paper are only 2 dimensionality, the knn parameters kr, kc we used to produce the results of the paper are 5&5 respectively:

**********code**********

```
rng(213);
```

```
% load the associate data file
load('Data/Small_CliTri.mat');

opt  = {'optimisation', 'ga', 'obj_option', 'quant_knn', 'obj_para',
{'knn', 5,5}};

para_range = {'eta1',[-1, 10],'eta2', [-1, 10],'alpha',[0, 3], 'beta',
[0,3]};

I    = {'R', R, 'method', 'LEDeig','dim',2, 'model_optimisation', opt,
'parameter_range',para_range };

led    = LEDF(I);
led    = training(led);


figure;
Plotcluster(led.X',lx,led.Y',ly);
```

**********code**********

Clinical Trial Text Data

We searched the results from 2 to 20 dimensions and the knn score parameters setting are 10&15.

**********code**********

```
rng(213);
% load the associate data file
load('Data/ClinicalTrial.mat');

led = cell(20,1);

for dim = 2 : 20
    opt  = {'optimisation', 'ga', 'obj_option', 'quant_knn', 'obj_para',
{'knn', 10,15}};

    para_range = {'eta1',[-1, 10],'eta2', [-1, 10],'alpha',[0, 3], 'beta',
[0,3]};

    I    = {'R', R, 'method', 'LEDeig','dim',dim, 'model_optimisation',
opt, 'parameter_range',para_range };

    led{dim-1}    = LEDF(I);
    led{dim-1}    = training(led{dim-1});
end
```

**********code**********

Reuters Text Data

We searched the results from 2 to 20 dimensions and the knn score parameters setting are 5&5. The co-embedding for the proposed method was produced through the following code:

**********code**********

```
rng(213);
% load the associate data file
```

```
load('Data/Reuters.mat');

led = cell(20,1);

for dim = 2 : 20
    opt   = {'optimisation', 'ga', 'obj_option', 'quant_knn', 'obj_para',
{'knn', 5,5}};

    para_range = {'eta1',[-1, 10],'eta2', [-1, 10],'alpha',[0, 3], 'beta',
[0,3]};

    I     = {'R', R, 'method', 'LEDeig','dim',dim, 'model_optimisation',
opt, 'parameter_range',para_range };

    led{dim-1}     = LEDF(I);
    led{dim-1}     = training(led{dim-1});
end
```

**********code**********

Cora Citation Data

The code for obtaining the Cora data embedding in 2 to 10 dimensional space is as follow:

**********code**********

```
rng(213);

% load the associate data file
load('Data/processed_cora.mat');

led = cell(10,1);

for dim = 2 : 10
    opt   = {'optimisation', 'ga', 'obj_option', 'quant_knn', 'obj_para',
{'knn', 5,5}};

    para_range = {'eta1',[-1, 10],'eta2', [-1, 10],'alpha',[0, 3], 'beta',
[0,3]};

    I     = {'R', R, 'method', 'LEDeig','dim',dim, 'model_optimisation',
opt, 'parameter_range',para_range };

    led{dim-1}     = LEDF(I);
    led{dim-1}     = training(led{dim-1});
end
```

**********code**********

Citeseer Citation Data

The code for obtaining the Citesser data embedding in 2 to 10 dimensional space is as follow:

**********code**********

```
rng(213);
% load the associate data file
```

```
load('Data/processed_citeseer.mat');

led = cell(10,1);

for dim = 2 : 10
    opt  = {'optimisation', 'ga', 'obj_option', 'quant_knn', 'obj_para',
{'knn', 5,5}};

    para_range = {'eta1',[-1, 10],'eta2', [-1, 10],'alpha',[0, 3], 'beta',
[0,3]};

    I    = {'R', R, 'method', 'LEDeig','dim',dim, 'model_optimisation',
opt, 'parameter_range',para_range };

    led{dim-1}    = LEDF(I);
    led{dim-1}    = training(led{dim-1});
end
**********************************code*********************************
```

Industry Company Sector Data

This dataset results are also produce under settings the knn score parameters as 5&5:

```
**********************************code*********************************
rng(213);

% load the associate data file
load('Data/processed_industry.mat');

led = cell(10,1);

for dim = 2 : 10
    opt  = {'optimisation', 'ga', 'obj_option', 'quant_knn', 'obj_para',
{'knn', 5,5}};

    para_range = {'eta1',[-1, 10],'eta2', [-1, 10],'alpha',[0, 3], 'beta',
[0,3]};

    I    = {'R', R, 'method', 'LEDeig','dim',dim, 'model_optimisation',
opt, 'parameter_range',para_range };

    led{dim-1}    = LEDF(I);
    led{dim-1}    = training(led{dim-1});
end

save('industry_LEDF.mat');

**********************************code*********************************
```