In [3]:

```
import numpy as np
import pyarrow.feather as feather
import pandas as pd
from surprise.model_selection import train_test_split, GridSearchCV, RandomizedSearc
from joblib import parallel_backend
from surprise import CoClustering
from sklearn.metrics import mean_squared_error
from surprise import Dataset, Reader, BaselineOnly, accuracy
```

In [2]:

```
df_train = feather.read_feather('netflix-5k.train.feather')
df_val = feather.read_feather('netflix-5k.validation.feather')
df_titles = feather.read_feather('netflix-5k.movie_titles.feather')
```

In [3]:

```
reader = Reader(rating_scale=(0, 5))
data = Dataset.load_from_df(df_train[['userID','movieID',
                                      'rating']],reader)
datav = Dataset.load_from_df(df_val[['userID','movieID',
                                     'rating']],reader)
trainset = data.build_full_trainset()
NA,valset = train_test_split(datav, test_size=1.0)
```

In [4]:

```
param_grid_svd = {'n_epochs': [10, 15, 20, 25, 30, 35],
                  'n_cltr_u': [1, 3, 5, 7, 9],
                  'n_cltr_i': [1, 3, 5, 7, 9]
                  }

with parallel_backend('multiprocessing', n_jobs=-1):
    gs_coclustering = RandomizedSearchCV(CoClustering, param_grid_svd, measures=['rm
                           n_jobs=-1, joblib_verbose=10)
    gs_coclustering.fit(data)

print("Best Score from Grid Search is ", gs_coclustering.best_score['rmse'])
print("Best parameters for CoCluserting are", gs_coclustering.best_params['rmse'])
```

```
[Parallel(n_jobs=-1)]: Using backend MultiprocessingBackend with 8 con
current workers.
[Parallel(n_jobs=-1)]: Done   2 tasks      | elapsed:   32.0s
[Parallel(n_jobs=-1)]: Done   9 tasks      | elapsed:  1.2min
[Parallel(n_jobs=-1)]: Done  16 tasks      | elapsed:  2.0min
[Parallel(n_jobs=-1)]: Done  25 tasks      | elapsed:  2.6min
[Parallel(n_jobs=-1)]: Done  30 out of  40 | elapsed:  3.5min remainin
g:  1.2min
[Parallel(n_jobs=-1)]: Done  35 out of  40 | elapsed:  3.7min remainin
g:   31.6s
[Parallel(n_jobs=-1)]: Done  40 out of  40 | elapsed:  3.8min remainin
g:    0.0s
[Parallel(n_jobs=-1)]: Done  40 out of  40 | elapsed:  3.8min finished

Best Score from Grid Search is  0.8827998834399162
Best parameters for CoCluserting are {'n_epochs': 35, 'n_cltr_u': 7,
'n_cltr_i': 9}
```

In [1]:

```
score = {'n_epochs': 35, 'n_cltr_u': 7, 'n_cltr_i': 9}
```

In [4]:

```
dataset = pd.DataFrame({'Parameters': score.keys(), 'Values': score.values()})
dataset
```

Out[4]:

	Parameters	Values
0	n_epochs	35
1	n_cltr_u	7
2	n_cltr_i	9