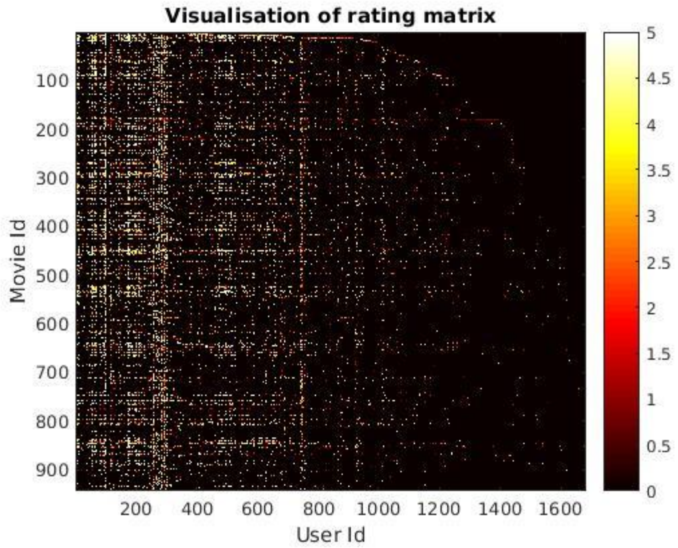
A survey of matrix completion and clustering techniques in the context of Recommender Systems

Niharika Shimona D’Souza

The challenge

- MovieLens 100k dataset:
- 100,000 ratings (1-5) ,943 users,1682 movies
- Highly sparse!

Task 1: Matrix Completion



- 19 genres : Action ,Adventure,Animation,Children's ,Comedy,Crime, Documentary,Drama,Fantasy, Film-Noir ,Horror,Musical,Mystery, Romance,Sci-Fi,Thriller,War,Western

Task 2: Genre Based Clustering

Datasets

Small Datasets:

- Comedy (505 movies)
- Thriller (251 movies)

Medium Dataset:

Comedy + Thriller (756 movies) : Minimise overlap of movies across genre!

Large Dataset:

Overlap of genre when all movies are considered (multilabel ?)

Datasets

Small Datasets:

- Comedy (505 movies)
- Thriller (251 movies)

Medium Dataset:

Comedy + Thriller (756 movies) : Minimise overlap of movies across genre!

Large Dataset:

Overlap of genre when all movies are considered (multilabel ?)

Part 1: Matrix Completion

Low Rank Matrix Completion via Proximal Gradient:

$$\min_{A} \quad \tau \|A\|_* + \frac{1}{2}\|A\|_F^2 \quad \text{s.t.} \quad \mathcal{P}_\Omega(A) = \mathcal{P}_\Omega(X),$$

Algorithm 3.2 (Low-Rank Matrix Completion by Proximal Gradient)

Input: Entries x_{ij} of a matrix $X \in \mathbb{R}^{D\times N}$ for $(i,j) \in \Omega$ and parameter $\tau > 0$.

Initialize $Z \leftarrow \mathbf{0}$.
repeat
    $A \leftarrow \mathcal{D}_\tau(\mathcal{P}_\Omega(Z))$.
    $Z \leftarrow Z + \beta(\mathcal{P}_\Omega(X) - \mathcal{P}_\Omega(A))$.
until convergence of Z.

Output: Matrix A.

Soft-Impute: [1]

$$\underset{Z}{\text{minimize}} \quad f_\lambda(Z) := \frac{1}{2}\|P_\Omega(X) - P_\Omega(Z)\|_F^2 + \lambda\|Z\|_* .$$

Algorithm 1 SOFT-IMPUTE

1. Initialize $Z^{\text{old}} = 0$.
2. Do for $\lambda_1 > \lambda_2 > \ldots > \lambda_K$:
   (a) Repeat:
      i. Compute $Z^{\text{new}} \leftarrow \mathbf{S}_{\lambda_k}(P_\Omega(X) + P_\Omega^\perp(Z^{\text{old}}))$.
      ii. If $\frac{\|Z^{\text{new}} - Z^{\text{old}}\|_F^2}{\|Z^{\text{old}}\|_F^2} < \varepsilon$ exit.
      iii. Assign $Z^{\text{old}} \leftarrow Z^{\text{new}}$.
   (b) Assign $\hat{Z}_{\lambda_k} \leftarrow Z^{\text{new}}$.
3. Output the sequence of solutions $\hat{Z}_{\lambda_1}, \ldots, \hat{Z}_{\lambda_K}$.

[1] Mazumder, Rahul, Trevor Hastie, and Robert Tibshirani. "Spectral regularization algorithms for learning large incomplete matrices." *Journal of machine learning research* 11.Aug (2010): 2287-2322.

Quantifying matrix completion performance

	Proximal Gradient	Soft-Impute
Small Dataset 1	1.8651 (rank = 91)	1.632 (rank = 93)
Small Dataset 2	1.93 (rank =66)	1.915 (rank =63)
Medium Dataset	2.479 (rank =153)	2.315 (rank =151)
Large Dataset	2.9015 (rank =202)	2.78 (rank = 204)

Table 1 : RMSE errors from 10 fold Cross Validation

$$\mathbf{RMSE}(X, \hat{A}) = \sqrt{\frac{1}{\Omega_s} \sum_{i,j \in \Omega_s} (X_{ij} - \hat{A}_{ij})}$$

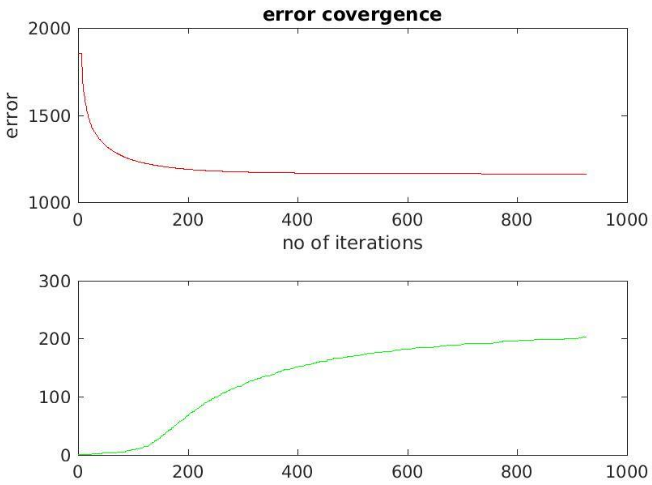


Figure 2: Performance run on a large dataset (Proximal Gradient)

Evaluation

- Soft-Impute gives better results
  - ➔ These completion results are used for clustering
- Performance on smaller of the two datasets is worse
  - ➔ Relatively larger percentage of missing entries
- Complete large matrix not considered for clustering (genre overlap)
  - ➔ 4 genres: Comedy, Documentary, Film-Noir,Thriller

Evaluation

- Soft-Impute gives better results
  - ➔ These completion results are used for clustering
- Performance on smaller of the two datasets is worse
  - ➔ Relatively larger percentage of missing entries
- Complete large matrix not considered for clustering (genre overlap)
  - ➔ 4 genres: Comedy, Documentary, Film-Noir,Thriller

Evaluation

- Soft-Impute gives better results
    - ➔ Completion results used for clustering
- Performance on smaller of the two datasets is worse
    - ➔ Relatively larger percentage of missing entries
- Complete large matrix not considered for clustering (genre overlap)
    - ➔ 4 genres: Comedy, Documentary, Film-Noir, Thriller

Part 2: Clustering Algorithms

1. K means:

$$\min_{\{\boldsymbol{\mu}_i\},\{w_{ij}\}} \sum_{i=1}^{n}\sum_{j=1}^{N} w_{ij}\|\boldsymbol{x}_j - \boldsymbol{\mu}_i\|^2$$

$$\text{s.t.} \quad w_{ij} \in \{0,1\} \text{ and } \sum_{i=1}^{n} w_{ij} = 1, \quad j = 1,\dots,N.$$

2. Spectral Clustering:

Affinity measures

- k-NN
- rbf

The silhouette coefficient

$$s_i = (b_i - a_i)/\max(a_i, b_i).$$

Algorithm 4.4 (K-means)

Input: A set of points $\{\boldsymbol{x}_j\}_{j=1}^N$ and the number of groups n.

1: **Initialization:** Select n distinct data points as initial cluster centers $\boldsymbol{\mu}_1^{(0)},\dots,\boldsymbol{\mu}_n^{(0)}$.

2: **while** (the clusters and their centers do not converge) **do**

3: Assign each data point $\boldsymbol{x}_j$ to its closest cluster center $\boldsymbol{\mu}_i^{(k)}$, i.e.,

$$w_{ij}^{(k+1)} \leftarrow \begin{cases} 1 & \text{if } i = \underset{\ell=1,\dots,n}{\arg\min}\|\boldsymbol{x}_j - \boldsymbol{\mu}_\ell^{(k)}\|_2^2, \\ 0 & \text{else.} \end{cases} \tag{4.55}$$

4: Update the cluster centers $\boldsymbol{\mu}_i^{(k+1)}$ to be the mean of all points $\boldsymbol{x}_j$ that belong to cluster i,

$$\boldsymbol{\mu}_i^{(k+1)} \leftarrow \frac{\sum_{j=1}^N w_{ij}^{(k+1)}\boldsymbol{x}_j}{\sum_{j=1}^N w_{ij}^{(k+1)}}. \tag{4.56}$$

If more than one cluster achieves the minimum, assign the point to one of them.

5: **end while**

Output: The n cluster centers $\{\boldsymbol{\mu}_i\}$ and the segmentation $\{w_{ij}\}$.

Algorithm 4.5 (Spectral Clustering)

Input: Number of clusters n and affinity matrix $W \in \mathbb{R}^{N\times N}$ for points $\{\boldsymbol{x}_j\}_{j=1}^N$.

1: Construct an affinity graph $\mathcal{G}$ with weight matrix W.

2: Compute the degree matrix $\mathcal{D} = \operatorname{diag}(W\mathbf{1})$ and the Laplacian $\mathcal{L} = \mathcal{D} - W$.

3: Compute the n eigenvectors of $\mathcal{L}$ associated with its n smallest eigenvalues.

4: Let $\boldsymbol{y}_1,\dots,\boldsymbol{y}_N$ be the columns of $Y \doteq [\boldsymbol{u}_1,\dots,\boldsymbol{u}_n]^\top \in \mathbb{R}^{n\times N}$, where $\{\boldsymbol{u}_i\}_{i=1}^n$ are the eigenvectors in step 3 normalized to unit Euclidean norm.

5: Cluster the points $\{\boldsymbol{y}_j\}_{j=1}^N$ into n groups using the K-means algorithm, Algorithm 4.4.

Output: The segmentation of the data into n groups.

Quantifying clustering performance

- Medium dataset

	Dataset 1	Dataset 2
K means	39.50	54.14
Spectral Clustering (k-NN)	13.07	49.79

Table 2: Percentage error on the datasets

Before separation	After separation
2.16	2.09

Table 3: Comparison of RMSE values before and after clustering

Fig 3:

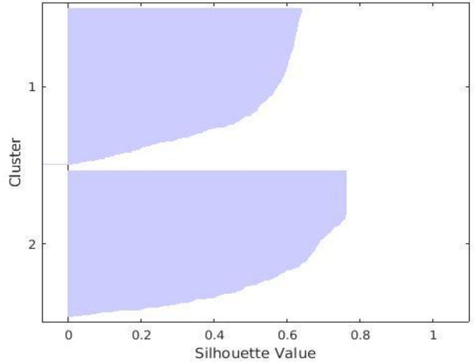


Fig 3: Silhouette coefficients on K Means clusters

Fig 4:

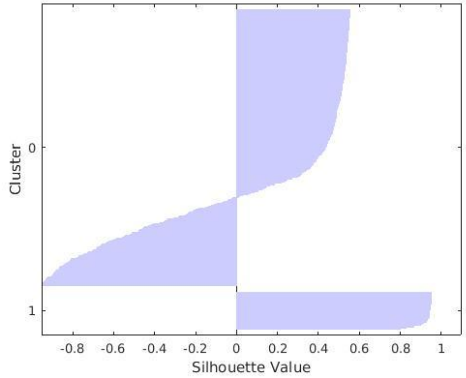


Fig 3: Silhouette coefficients on clusters from Spectral Clustering

Quantifying clustering performance

- Large dataset

	Dataset 1	Dataset 2	Dataset 3	Dataset 4
K means	56.56	34.79	67.27	46.98
Spectral Clustering (k-NN)	53.34	36.67	65.78	44.05

Table 4: Percentage error on the datasets

Fig 5:

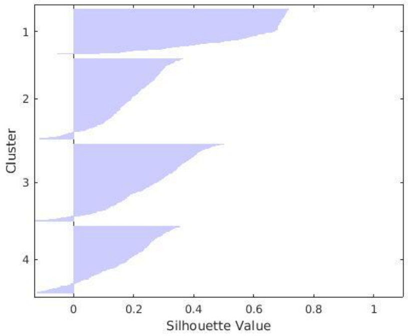
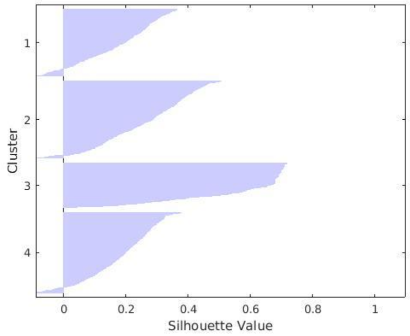


Fig 5: Silhouette coefficients on K Means clusters

Fig 6: Silhouette coefficients on clusters from Spectral Clustering

Fig 6:



Discussions and Future work

- Overall clustering performance is poor $\rightarrow$ matrix completion isn't good enough.
- Outliers in the data, corrupted entries not accounted for.

Robustness:
$$\min_{L,E} \quad \|L\|_* + \lambda\|E\|_1 \quad \text{s.t.} \quad \mathcal{P}_\Omega(X) = \mathcal{P}_\Omega(L+E).$$

- Better matrix completion techniques to be explored

Matrix Factorisation:
$$\min_{U,V,d} \|P_\Omega(X - UV^\top)\|_F^2 + \lambda_u\|U\|_F^2 + \lambda_v\|V\|_F^2,$$

References

Vidal, Rene, Yi Ma, and Shankar Sastry. "Generalized principal component analysis (GPCA)." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27.12 (2005): 1945-1959.

Mazumder, Rahul, Trevor Hastie, and Robert Tibshirani. "Spectral regularization algorithms for learning large incomplete matrices." *Journal of machine learning research* 11.Aug (2010): 2287-2322.

Srebro, Nathan, Jason Rennie, and Tommi S. Jaakkola. "Maximum-margin matrix factorization." *Advances in neural information processing systems*. 2005.

Hastie, Trevor, et al. "Matrix completion and low-rank svd via fast alternating least squares." *J. Mach. Learn. Res* 16.1 (2015): 3367-3402.

Thank You