Semantics in Society

Oscar I. Hernandez

Presentation for the ICS691C Seminar: Language Models and Computation

March 1, 2024

oih@hawaii.edu
University of Hawai'i at Mānoa, College of Natural Sciences

Netflix Problem from the ICS691C lecture notes [1]

L	Interstellar	Juno	Kagemusha	Legend
Alice	⋆⋆⋆⋆	⋆⋆	⋆⋆⋆⋆	⋆
Bob	⋆	⋆⋆⋆	⋆⋆	
Carol	⋆⋆	⋆	⋆⋆⋆⋆⋆	
Dave	⋆	⋆⋆⋆⋆	⋆⋆⋆	
Ed		⋆⋆	⋆⋆⋆⋆⋆	⋆⋆

Table: The raw star rating induces the normalized rating [1]

Problem: The matrix is missing entries, which are to be predicted.

Netflix Problem from the ICS691C lecture notes [1]

L	Interstellar	Juno	Kagemusha	Legend
Alice	⋆⋆⋆⋆	⋆⋆	⋆⋆⋆⋆	⋆
Bob	⋆	⋆⋆⋆	⋆⋆	
Carol	⋆⋆	⋆	⋆⋆⋆⋆⋆	
Dave	⋆	⋆⋆⋆⋆	⋆⋆⋆	
Ed		⋆⋆	⋆⋆⋆⋆⋆	⋆⋆

L	Interstellar	Juno	Kagemusha	Legend
Alice	1.25	0.83	0	-0.12
Bob	1.05	1.13	0.35	
Carol	1.12	1.02	0.21	
Dave	1.57	0.35	-0.56	
Ed		0.18	1.02	0.98

Table: The raw star rating induces the normalized rating [1]

Problem: The matrix is missing entries, which are to be predicted.

Netflix Problem from the ICS691C lecture notes [1]

L	Interstellar	Juno	Kagemusha	Legend
Alice	$\star\star\star\star$	$\star\star$	$\star\star\star\star$	$\star$
Bob	$\star$	$\star\star\star$	$\star\star$	
Carol	$\star\star$	$\star$	$\star\star\star\star\star$	
Dave	$\star$	$\star\star\star\star$	$\star\star\star$	
Ed		$\star\star$	$\star\star\star\star\star$	$\star\star$

L	Interstellar	Juno	Kagemusha	Legend
Alice	1.25	0.83	0	-0.12
Bob	1.05	1.13	0.35	
Carol	1.12	1.02	0.21	
Dave	1.57	0.35	-0.56	
Ed		0.18	1.02	0.98

Table: The raw star rating induces the normalized rating [1]

Problem: The matrix is missing entries, which are to be predicted.
Solution: Latent Semantic Analysis

Outline

(Conjugate) Transpose of $A = (a_{ij})_{(m,n)} : \mathbb{C}^n \to \mathbb{C}^m$

L	Interstellar	Juno	Kagemusha	Legend
Alice	1.25	0.83	0	-0.12
Bob	1.05	1.13	0.35	
Carol	1.12	1.02	0.21	
Dave	1.57	0.35	-0.56	
Ed		0.18	1.02	0.98

“Reflect“ matrix A so its columns are rows of $A^* = (a_{ji})_{(n,m)} : \mathbb{C}^m \to \mathbb{C}^n$

$$\begin{pmatrix} 1.25 & 0.83 & 0 & -0.12 \\ 1.05 & 1.13 & 0.35 & \\ 1.12 & 1.02 & 0.21 & \\ 1.57 & 0.35 & -0.56 & \\ & 0.18 & 1.012 & 0.98 \end{pmatrix}^* = \begin{pmatrix} 1.25 & 1.05 & 1.12 & 1.57 & \\ 0.83 & 1.13 & 1.02 & 0.35 & 0.18 \\ 0 & 0.35 & 0.21 & -0.56 & 1.02 \\ -0.12 & & & & 0.98 \end{pmatrix} \quad (1)$$

(replacing each entry $z = a + ib$ with its complex conjugate $\overline{z} = a - ib$).

Singular Value Decomposition

Theorem 0 (Singular Value Decomposition (S. V. D.))

The matrix $A : \mathbb{C}^n \to \mathbb{C}^m$ *has an SVD, i.e. a triple* $\{U, \Sigma, V\}$ *satisfying*

$$L = U \cdot \Sigma \cdot V^* \tag{2}$$

where U, V *are unitary matrices satisfying* $U^*U = UU^* = I$, $V^*V = I$, *and* Σ *is a diagonal matrix whose entries are real.*

$$\begin{pmatrix} 1.25 & 1.05 & 1.12 & 1.57 \\ 0.83 & 1.13 & 1.02 & 0.35 \\ 0 & 0.35 & 0.21 & -0.56 \end{pmatrix} = \begin{pmatrix} 0.83 & -0.4 \\ 0.55 & 0.6 \\ 0 & 0.7 \end{pmatrix} \cdot \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0.5 & 0.5 & 0.5 & 0.5 \\ 0 & 0.5 & 0.3 & -0.8 \end{pmatrix} \tag{3}$$

Singular Value Decomposition

Theorem 0 (Singular Value Decomposition (S. V. D.))

The matrix $A : \mathbb{C}^n \to \mathbb{C}^m$ has an SVD, i.e. a triple $\{U, \Sigma, V\}$ satisfying

$$L = U \cdot \Sigma \cdot V^* \tag{2}$$

where U, V are unitary matrices satisfying $U^*U = UU^* = I$, $V^*V = I$, and Σ is a diagonal matrix whose entries are real.

$$\begin{pmatrix} 1.25 & 1.05 & 1.12 & 1.57 \\ 0.83 & 1.13 & 1.02 & 0.35 \\ 0 & 0.35 & 0.21 & -0.56 \end{pmatrix} = \begin{pmatrix} 0.83 & -0.4 \\ 0.55 & 0.6 \\ 0 & 0.7 \end{pmatrix} \cdot \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0.5 & 0.5 & 0.5 & 0.5 \\ 0 & 0.5 & 0.3 & -0.8 \end{pmatrix} \tag{3}$$

$$\begin{pmatrix} 1.25 & 0.83 & 0 \\ 1.05 & 1.13 & 0.35 \\ 1.12 & 1.02 & 0.21 \\ 1.57 & 0.35 & -0.56 \end{pmatrix} = \begin{pmatrix} -0.5 & 0 & 0.31 \\ -0.5 & -0.5 & 0.5 \\ -0.5 & -0.31 & -0.81 \\ -0.5 & 0.81 & 0 \end{pmatrix} \begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0.01 \end{pmatrix} \begin{pmatrix} -0.83 & -0.56 & 0 \\ 0.4 & -0.6 & -0.69 \\ 0.39 & -0.58 & 0.72 \end{pmatrix} \tag{4}$$

Singular Value Decomposition

Theorem 0 (Singular Value Decomposition (S. V. D.))

The matrix $A : \mathbb{C}^n \to \mathbb{C}^m$ has an SVD, i.e. a triple $\{U, \Sigma, V\}$ satisfying

$$L = U \cdot \Sigma \cdot V^* \tag{2}$$

where U, V are unitary matrices satisfying $U^*U = UU^* = I$, $V^*V = I$, and Σ is a diagonal matrix whose entries are real.

$$\begin{pmatrix} 1.25 & 1.05 & 1.12 & 1.57 \\ 0.83 & 1.13 & 1.02 & 0.35 \\ 0 & 0.35 & 0.21 & -0.56 \end{pmatrix} = \begin{pmatrix} 0.83 & -0.4 \\ 0.55 & 0.6 \\ 0 & 0.7 \end{pmatrix} \cdot \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0.5 & 0.5 & 0.5 & 0.5 \\ 0 & 0.5 & 0.3 & -0.8 \end{pmatrix} \tag{3}$$

$$\begin{pmatrix} 1.25 & 0.83 & 0 \\ 1.05 & 1.13 & 0.35 \\ 1.12 & 1.02 & 0.21 \\ 1.57 & 0.35 & -0.56 \end{pmatrix} = \begin{pmatrix} -0.5 & 0 & 0.31 \\ -0.5 & -0.5 & 0.5 \\ -0.5 & -0.31 & -0.81 \\ -0.5 & 0.81 & 0 \end{pmatrix} \begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0.01 \end{pmatrix} \begin{pmatrix} -0.83 & -0.56 & 0 \\ 0.4 & -0.6 & -0.69 \\ 0.39 & -0.58 & 0.72 \end{pmatrix} \tag{4}$$

The (conjugate) transpose $L^* : \mathbb{C}^m \to \mathbb{C}^n$ has the following SVD.

$$L^* = (U\Sigma V^*)^* = (V^*)^* \Sigma^* U^* = V\Sigma U^* \tag{5}$$

Singular Value Decomposition

The compositions $L^*L : \mathbb{C}^n \to \mathbb{C}^n$, $LL^* : \mathbb{C}^m \to \mathbb{C}^m$ defined by

$$L^*L = V\Sigma\left(U^*U\right)\Sigma V^* = V\Sigma^2 V^* \tag{6}$$
$$LL^* = U\Sigma\left(V^*V\right)\Sigma U^* = U\Sigma^2 U^* \tag{7}$$

are such that

1. they are symmetric;

$$\left(L^*L\right)^* = L^*\left(L^*\right)^* = L^*L \tag{8}$$
$$\left(LL^*\right)^* = \left(L^*\right)^* L^* = LL^*; \tag{9}$$

2. the eigenvalues of both are the non-zero entries on the diagonal of Σ^2;
3. the eigenvectors of LL^* are the columns of U;
4. the eigenvectors of L^*L are the columns of V, and we call them the right singular vectors of L.

Question

1. How could one predict the missing entries in an incomplete matrix

$$\begin{pmatrix} 1.25 & 0.83 & 0 \\ 1.05 & 1.13 & 0.35 \\ 1.12 & ? & 0.21 \\ 1.57 & 0.35 & -0.56 \end{pmatrix} \tag{10}$$

using data from $\begin{pmatrix} 1.12 & ? & 0.21 \end{pmatrix}$ and the maximal complete matrix

$$\begin{pmatrix} 1.25 & 0.83 & 0 \\ 1.05 & 1.13 & 0.35 \\ 1.57 & 0.35 & -0.56 \end{pmatrix}? \tag{11}$$

Question

1. How could one predict the missing entries in an incomplete matrix

$$\begin{pmatrix} 1.25 & 0.83 & 0 \\ 1.05 & 1.13 & 0.35 \\ 1.12 & ? & 0.21 \\ 1.57 & 0.35 & -0.56 \end{pmatrix} \tag{10}$$

   using data from $\begin{pmatrix} 1.12 & ? & 0.21 \end{pmatrix}$ and the maximal complete matrix

$$\begin{pmatrix} 1.25 & 0.83 & 0 \\ 1.05 & 1.13 & 0.35 \\ 1.57 & 0.35 & -0.56 \end{pmatrix}? \tag{11}$$

2. How would the prediction depend on the other users and their concepts?

Idea

1. Compute an SVD of the maximal complete matrix:

$$\begin{pmatrix} 1.25 & 0.83 & 0 \\ 1.05 & 1.13 & 0.35 \\ 1.57 & 0.35 & -0.56 \end{pmatrix} = U\Sigma V^*. \tag{12}$$

Idea

1. Compute an SVD of the maximal complete matrix:

$$\begin{pmatrix} 1.25 & 0.83 & 0 \\ 1.05 & 1.13 & 0.35 \\ 1.57 & 0.35 & -0.56 \end{pmatrix} = U\Sigma V^*. \tag{12}$$

2. Identify the index j of the missing column in the incomplete row.

$$u = \begin{pmatrix} 1.02 & ? & 0.21 \end{pmatrix} \tag{13}$$

Idea

1. Compute an SVD of the maximal complete matrix:

$$\begin{pmatrix} 1.25 & 0.83 & 0 \\ 1.05 & 1.13 & 0.35 \\ 1.57 & 0.35 & -0.56 \end{pmatrix} = U\Sigma V^*. \tag{12}$$

2. Identify the index j of the missing column in the incomplete row.

$$u = \begin{pmatrix} 1.02 & ? & 0.21 \end{pmatrix} \tag{13}$$

3. Obtain u_1, V_1^* by removing column j from u and V^*.

Idea

1. Compute an SVD of the maximal complete matrix:

$$\begin{pmatrix} 1.25 & 0.83 & 0 \\ 1.05 & 1.13 & 0.35 \\ 1.57 & 0.35 & -0.56 \end{pmatrix} = U\Sigma V^*. \tag{12}$$

2. Identify the index j of the missing column in the incomplete row.

$$u = \begin{pmatrix} 1.02 & ? & 0.21 \end{pmatrix} \tag{13}$$

3. Obtain u_1, V_1^* by removing column j from u and V^*.
4. Obtain u_1 as a linear combination of columns in V_1^*

$$u_1 = cV_1^* \tag{14}$$

by computing the projection $c = u_1 \left(V_1^*\right)^*$ of u_1 onto V_1^*.

Idea

1. Compute an SVD of the maximal complete matrix:

$$\begin{pmatrix} 1.25 & 0.83 & 0 \\ 1.05 & 1.13 & 0.35 \\ 1.57 & 0.35 & -0.56 \end{pmatrix} = U\Sigma V^*. \tag{12}$$

2. Identify the index j of the missing column in the incomplete row.

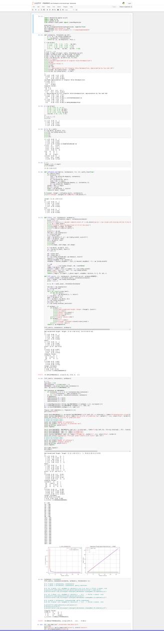
$$u = \begin{pmatrix} 1.02 & ? & 0.21 \end{pmatrix} \tag{13}$$

3. Obtain u_1, V_1^* by removing column j from u and V^*.
4. Obtain u_1 as a linear combination of columns in V_1^*

$$u_1 = cV_1^* \tag{14}$$

by computing the projection $c = u_1 \left(V_1^*\right)^*$ of u_1 onto V_1^*.
5. Compute cV^* as an estimate for the direction of u and, finally, approximate the entries of $\frac{u}{||u||}$ by $\frac{cV^*}{||cV^*||}$.

Demo in Jupyter Notebook



Vital Statistics on Congress, Brookings

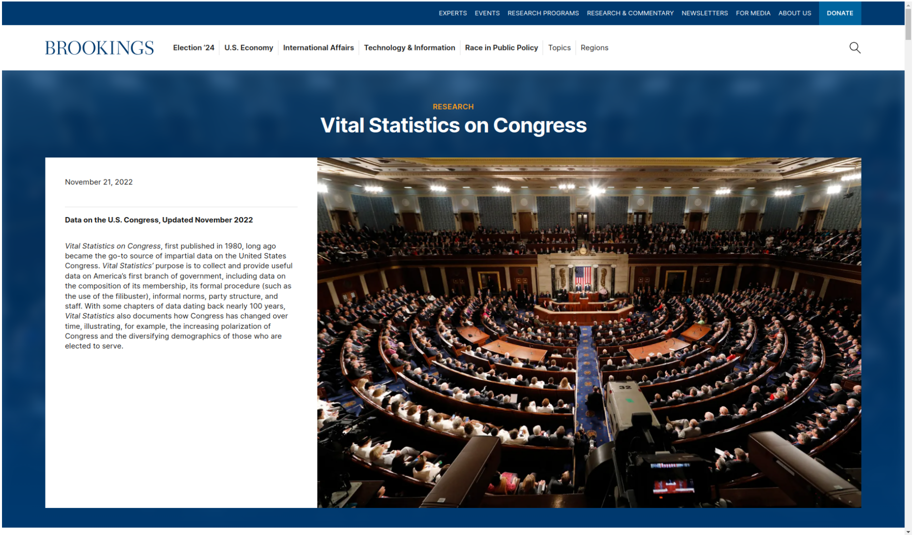


Figure: Party Unity Scores in Congressional Voting, 1945-2016 (percent) [2]

Social Media

A lonely girl in Washington State — a volunteer Sunday school teacher and part-time babysitter ...

— Singer & Brooking [3]

Social Media

A lonely girl in Washington State — a volunteer Sunday school teacher and part-time babysitter — described how ISIS recruiters gave her the attentive friends she'd always craved. (Only a sharp-eyed grandmother stopped her from boarding a plane to Syria.) ISIS promised adventure and a sense of belonging. "It's a closed community — almost a clique," explained terrorism analyst Seamus Hughes. "They share memes and inside jokes, terms and phrases you'd only know if you were a follower."

In each case, recruits to extremist causes are lured by a warmth and camaraderie that seems lacking in their own lonely lives. In each case, such recruits build communities that attract people from across the world but that show almost no diversity of thought.

— *LikeWar: The Weaponization of Social Media* by Singer & Brooking [3]

References

[1] Dusko Pavlovic. Semantics: The meaning of language. 2024.

[2] Molly E. Reynolds et al. Vital statistics on congress. `https://www.brookings.edu/articles/vital-statistics-on-congress`, November 2022. Accessed: 2024-02-23.

[3] P.W. Singer and E.T. Brooking. *LikeWar: The Weaponization of Social Media*. Houghton Mifflin Harcourt, 2018.