Dimensionality reduction & matrix factorization

Friday, 23 April 2021 08:18

Interaction matrix

$$R = [r_{ui}] = \underbrace{\left.\begin{bmatrix} \ddots & \vdots & \iddots \\ \cdots & r_{ui} & \cdots \\ \iddots & \vdots & \ddots \end{bmatrix}\right\} m}_{n}$$

$R \in M_{m \times n}$

m users

n items

User vectors (rows)

e.g. $r_u = \underbrace{[0, 1, 1, 0, 0, \ldots]}_{n}$

Item vectors (columns)

$r_i = \underbrace{[1, 0, 0, 0, 1, \ldots]}_{m}$

General model

$$\hat{r}_{ui} = f(r_u, r_i \mid \theta)$$

f - function dependent on parameters θ which has to be fit to data so that

$$\text{error} = \sum_u \sum_i d(r_{ui}, \hat{r}_{ui})$$

is minimized

The squared distance (as in RMSE or Euclidean distance)

$$d(r_{ui}, \hat{r}_{ui}) = (r_{ui} - \hat{r}_{ui})^2$$

is the most often used distance

In many applications most interactions between users and items are not known. Then only the error on the known

In many applications most interactions between users and items are not known. Then only the error on the known interactions is minimized.

Problem

p_u and q_i are very long and sparse (contain mostly zeros)

Solution

Reduce the dimensionality of user and item representations

$$r_u \in \mathbb{R}^n \rightarrow p_u \in \mathbb{R}^d \qquad r_i \in \mathbb{R}^m \rightarrow q_i \in \mathbb{R}^d$$

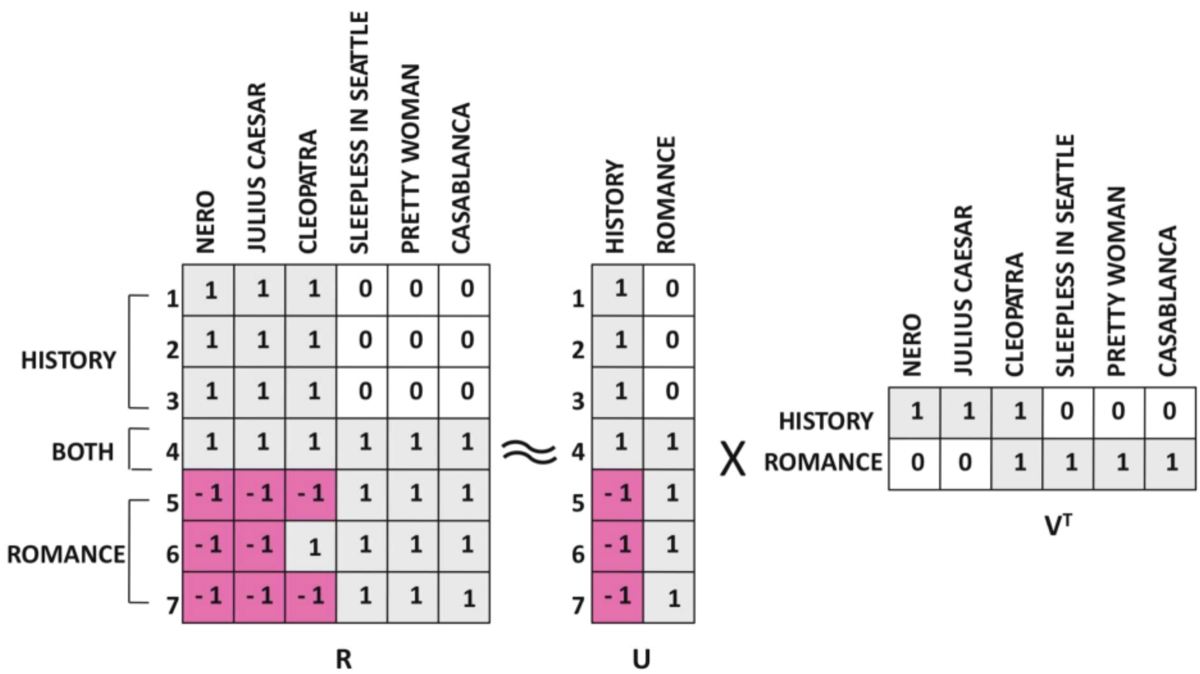
where $d \ll n, m$ (d much smaller than n and m)

How to do that?

- Dimensionality reduction (PCA, tSNE, UMAP)
- Matrix factorization

Matrix factorization

Example



Source: Aggarwal

Theorem (Singular Value Decomposition)

For every matrix $R \in M_{m\times n}$ there exist matrices $P \in M_{m\times m}$, $\Sigma \in M_{m\times n}$, $Q \in M_{n\times n}$ such that

$$R = P \Sigma Q^T$$

and

- rows of P are orthonormal eigenvectors of RR^T
- rows of Q are orthonormal eigenvectors of $R^T R$
- Σ is diagonal and the diagonal consists of square roots of all non-zero eigenvalues of RR^T (which are also all non-zero eigenvalues of $R^T R$)

A pair of eigenvector v and eigenvalue λ for a matrix A satisfy the following equation

$$A v = \lambda v$$

$$m\left\{\underbrace{\begin{bmatrix} & & \\ & r_{ui} & \\ & & \end{bmatrix}}_{n}\right. = m\left\{\underbrace{\begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix}}_{m}\right. \cdot\; m\left\{\underbrace{\begin{bmatrix} e_1 & & 0 \\ & e_2 & \\ 0 & & \ddots \end{bmatrix}}_{n}\right. \cdot\; n\left\{\underbrace{\begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}}_{n}\right.^{T}$$

$v_1, \dots, v_m$ form orthonormal basis of $\mathbb{R}^m$

... form orthonormal basis of $\mathbb{R}^n$

$v_1, \dots, v_m$ form ...

$w_1, \dots, w_n$ form orthonormal basis of $\mathbb{R}^n$

After changing notation to

$$\begin{bmatrix} p_1 \\ \vdots \\ p_m \end{bmatrix} = \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix} \begin{bmatrix} e_1 & & 0 \\ & e_2 & \\ 0 & & \ddots \end{bmatrix}$$

and

$$\begin{bmatrix} q_1 \\ \vdots \\ q_n \end{bmatrix} = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}$$

we have

$$\begin{bmatrix} & & \\ & r_{ui} & \\ & & \end{bmatrix} = \begin{bmatrix} p_1 \\ \vdots \\ p_m \end{bmatrix} \begin{bmatrix} q_1 \\ \vdots \\ q_n \end{bmatrix}^T$$

$$= \begin{bmatrix} p_1 q_1 & p_1 q_2 & \dots & p_1 q_n \\ p_2 q_1 & \dots & & \vdots \\ \vdots & & \vdots & \vdots \\ p_m q_1 & \dots & & p_m q_n \end{bmatrix}$$

$$\boxed{\forall_{u,i} \quad r_{ui} = p_u q_i}$$

Problem: $p_u \in \mathbb{R}^n$, $q_i \in \mathbb{R}^n$ where n is large

Solution: Approximate matrix R with only d largest eigenvalues, d first columns in P, d first columns in Q

d largest eigenvalues, d first columns in P, d first columns in Q

$$\Omega \approx P_d \Sigma_d Q_d^T$$

where

$$P_d \in M_{m \times d}, \quad \Sigma_d \in M_{d \times d}, \quad Q_d \in M_{n \times d}$$

Then

$$\forall_{u,i} \quad r_{ui} \approx p_u q_i$$

and $p_u \in \mathbb{R}^d$, $q_i \in \mathbb{R}^d$

(the higher the value of d, the more accurate this approximation is; but typically relatively very small d is enough to obtain very high accuracy)

Idea for a recommender

Find dense representation vectors $p_u \in \mathbb{R}^d$, $q_i \in \mathbb{R}^d$ such that

$$MSE = \frac{1}{|R|} \sum_{u,i} (r_{u,i} - \hat{r}_{u,i})^2 = \frac{1}{|R|} \sum_{u,i} (r_{u,i} - p_u \cdot q_i)^2$$

is minimized.

Here $|R|$ is the number of interactions used for training.

Then our model is given by

$$\hat{r}_{u,i} = f(r_u, r_i) = p_u \cdot q_i$$

and the recommender returns items i with the highest scores $\hat{r}$ for the given user

It can be proven that for $d = n$ minimizing the squared error as defined above yields exactly the same matrix ... as given by the Singular Value Decomposition

It can be ... error as defined above yields exactly the same matrix decomposition as given by the Singular Value Decomposition theorem!

MSE error can be minimized using many methods:

- SGD (Stochastic Gradient Descent)
- ALS (Alternating Least Squares)
- black box optimizers, e.g. Tree Parzen Estimator