Generation Meets Recommendation: Proposing Novel Items for Groups of Users
Vinh Vo Thanh et al. (2018)

June Kim

2019.01.15

Contents

Contents

Recommender Systems for Makers

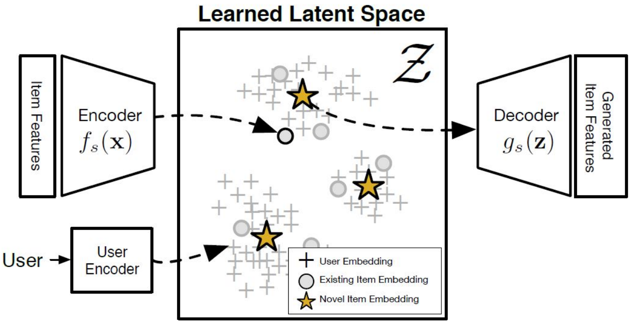
Proposing **new items** that could be created to satisfy users

For item makers, important questions include:

- **What set of products** would appeal to a wide range of people?
- **Who** would like those products?
- **How many** products should you make?

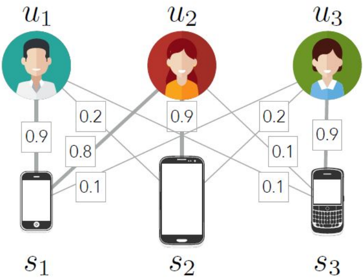
Overview

- Shared latent space learned from user-item rating data
- Generating plausible K items that appeal to different groups of users
- Generating item features via the decoder.



Contents

Problem Formulation



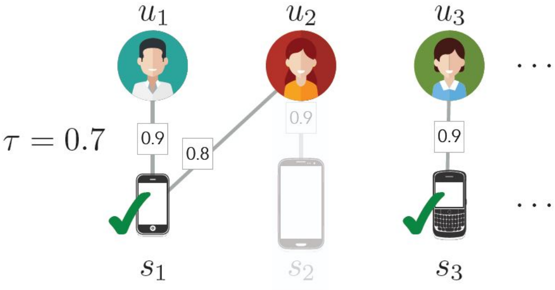
$\ldots \in \mathcal{U}$ (Set of Users)

$r(u, s)$ (Rating Function)

$\ldots \in \mathcal{S}$ (Set of Items)

"Cover" $$c(u, s) = \begin{cases} 1 & \text{if } \; r(u, s) > \tau \quad \text{(above threshold)} \\ 0 & \text{otherwise} \end{cases}$$

Problem Formulation



$\dots \in \mathcal{U}$ (Set of Users)

$r(u, s)$ (Rating Function)

$\dots \in \mathcal{S}$ (Set of Items)

$$\operatorname{argmax}_{\mathcal{V} \subseteq \mathcal{S}} \sum_{u \in \mathcal{U}} \sum_{v \in \mathcal{V}} r(u, v) c_{uv}$$

such that:

$$\sum_{v \in \mathcal{V}} c_{uv} = 1 \text{ and } |\mathcal{V}| = k$$

a user is only covered by one item

k items selected

Re-Formalizing with a Latent Space

The Maximum Coverage Novel Itemset Problem

For given $\hat{r}(z_u, z_s)$ and $\mathcal{D} = \{d^{(k)} = (u, i, x_i, r_{ui})^{(k)}\}_{k=1}^{|D|}$, seek the latent representation z_v and decoded feature $x_v = g_s(z_v)$ for novel covering items $v \in \mathcal{V}$ which satisfying :

$$\text{argmax}_{\mathcal{V}} \sum_{u \in \mathcal{U}} \sum_{v \in \mathcal{V}} \hat{r}(z_u, z_v) c_{uv}$$

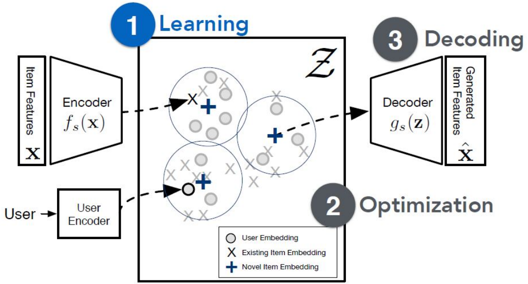
such that $\sum_{v \in \mathcal{V}} c_{uv} = 1$ and $|\mathcal{V}| = k$

- $z_s \in \mathcal{Z} \subseteq \mathbb{R}^{d_z}$: latent vector of item $s \in \mathcal{S}$
- $x_s \in \mathcal{X} \subseteq \mathbb{R}^{d_x}$: observed feature of item
- $z_u \in \mathcal{Z}$: latent vector of user
- $\hat{r}(z_u, z_s)$: pre-specified rating function (e.g. $z_u^T z_s$)

Contents

Learning the Latent Space

- Learning $\mathcal{Z}$ given $\hat{r}(z_u, z_s)$ and $\mathcal{D}$
- Learning encoder $f_s(x;\phi) : \mathcal{X} \rightarrow \mathcal{Z}$ and decoder $g_s(z;\theta) : \mathcal{Z} \rightarrow \mathcal{X}$
- Assume $\hat{r}(z_u, z_v) = f_r(z_u^T, z_s)$



$$\mathcal{L} = \text{Rating Loss} + \text{Reconstruction Loss} + \text{Regularizer}$$
$$= \mathcal{L}_r(\theta) + \lambda_1 \mathcal{L}_g(\phi) + \lambda_2 \mathcal{R}(\theta, \phi)$$

Linear Model

Assume both f_s and g_s are linear functions

$$z_s = f_s(x_s) = M_\theta x_s$$
$$x_s = g_s(z_s) = M_\phi z_s$$

Rating loss:

$$\mathcal{L}_r(\theta) = -\sum_{d \in \mathcal{D}} r_{ui} \log \tilde{r}_{ui} + (1 - r_{ui}) \log (1 - \tilde{r}_{ui})$$

$$(\tilde{r}_{ui} = \hat{r}(z_u, z_i) = \sigma(z_u^T, z_i) = \sigma(z_u^T M_\theta x_i))$$

Reconstruction loss:

$$\mathcal{L}_g(\phi) = -\sum_{d \in \mathcal{D}} \sum_j x_{ij} \log \tilde{x}_{ij} + (1 - x_{ij}) \log (1 - \tilde{x}_{ij})$$

$$(\tilde{x}_i = g_s(z_i))$$

Regularizer:

$$\mathcal{R}(\theta, \phi) = ||M_\theta||_F + ||M_\phi||_F$$

VAE

- Want to find parameter $\theta = \text{argmax}_\theta \mathbb{E}_{data}[\log p_\theta(d)]$
- Since $\log p_\theta(d) = \mathbb{E}_{p(z_i)}[p(r_{ui}|z_u, z_i)p_\theta(x_i|g_s(z_i))]$ is intractable, introduce $q_\phi(z_i|x_i)$ and perform inference by maximizing ELBO

$$\begin{aligned}\log p_\theta(d) \geq \ell(\theta, \phi; d) = {} & \mathbb{E}_{q(z_i|f_\phi(x_i))}[\log p(r_{ui}|z_u, z_i)] && \textbf{(Rating loss)} \\ & + \mathbb{E}_{q(z_i|f_\phi(x_i))}[\log p(x_i|z_i)] && \textbf{(Reconstruction loss)} \\ & - \mathbb{D}_{KL}[q_\phi(z_i|x_i)||p(z_i)] && \textbf{(Regularizer)}\end{aligned}$$

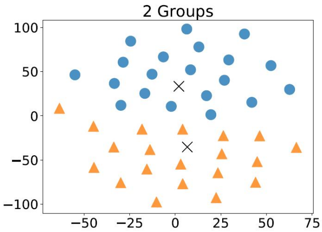
$$q_\phi(z_i|x_i) = \mathcal{N}(f_s^\mu(x_i), \text{diag}(\exp(f_s^\sigma(x_i))))$$
$$\log p(x_i|z_i) = \sum x_{ij} \log \tilde{x}_{ij} + (1 - x_{ij}) \log (1 - \tilde{x}_{ij})$$

Latent Space Searching

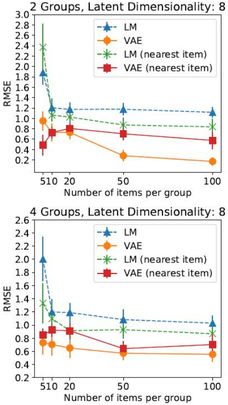
- Using greedy algorithm to find novel items that possess maximum coverage
  1. Discretize $\mathcal{Z}$ to obtain candidate items:
     sample T candidate points from $p(z_s|\mathcal{Z}_{\mathcal{I}})$
     train a Gaussian Mixture Model to learn $\mathcal{Z}_{\mathcal{I}}$
     $\mathcal{Z}_{\mathcal{I}}$ is the latent representation of the observed items
  2. successively pick the candidate item that attains the maximum total rating given by uncovered users until K samples are selected

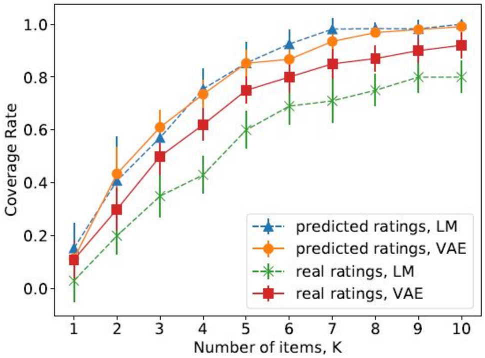
Contents

- Construct synthetic data with 5000 users split into evenly among K groups
- Sample N_g observed items $x_i \in \mathbb{R}^{20}$ from MN centered at ideal item for that group
- Observed items were rated differently depending on whether the user was from the same group or otherwise

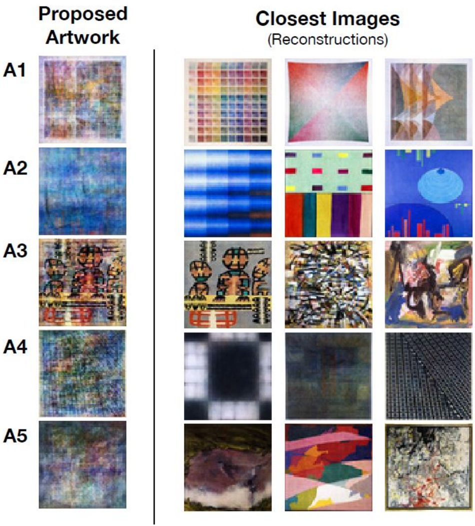


Quantitative Results



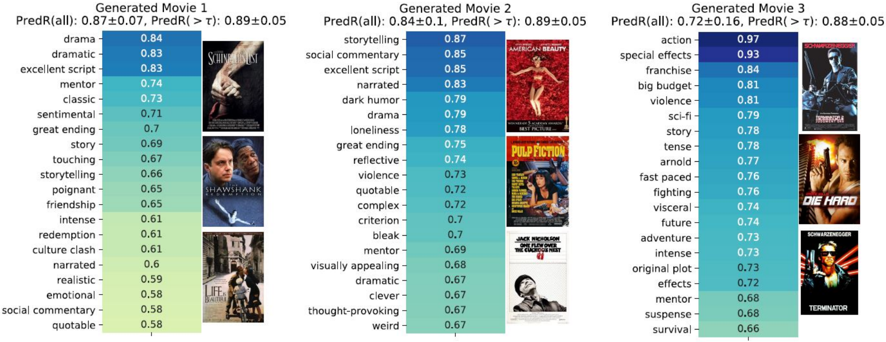


Qualitative Results: Generating Abstract Art



MART dataset comprises 500 art pieces by professional artists and 10 thousand ratings by 100 people

Qualitative Results: Generating Movielens Tag Genomes



Movielens Tag Genome and MovieLens-1M datasets contains 3952 movies, 6040 users and 1 million ratings. Each genome describes a given movie via 1128 tags with relevance scores between [0, 1]