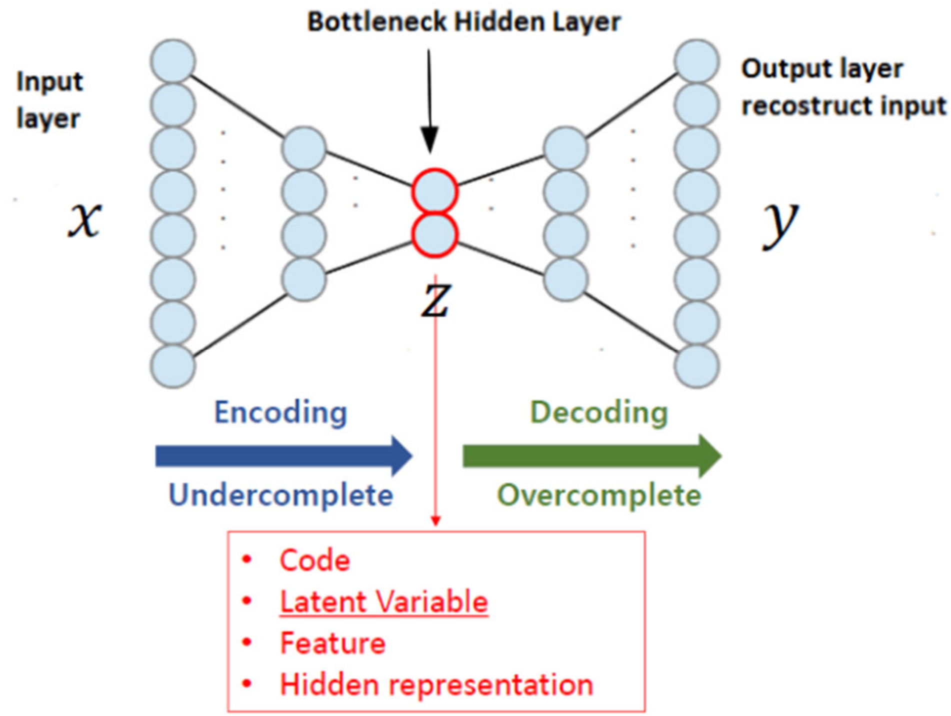
AutoRec & ConvMF

SoC 19, SeungAn Jung

AutoRec - Intro

- Autoencoder paradigm into CF model
- AutoRec has representational and computational advantages over existing neural approaches to CF
- Outperforms SOTA CF
  - RBM-CF
    - Generative Model - Discriminative Model
    - MLE - RMSE
    - Only applicable for discrete ratings - Agnostic to rating
  - LLORMA
    - Embed both users and items into shared latent space – only users/itmes
    - Linear latent representation – nonlinear latent representation

AutoRec - AutoEncoder



AutoRec - Model

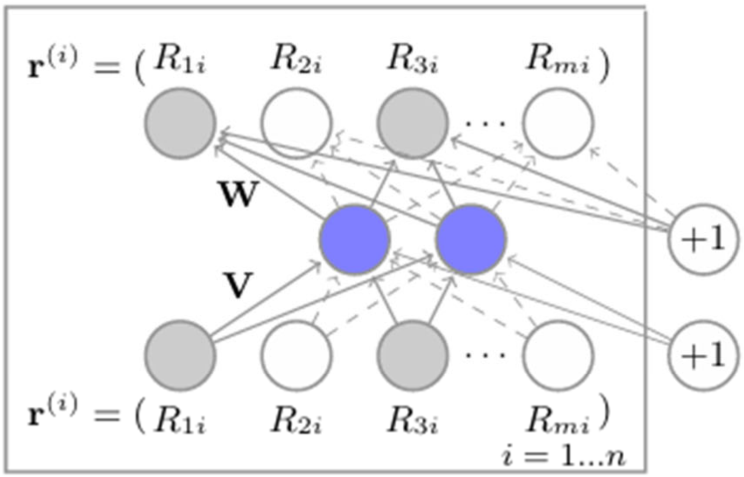


Figure 1: Item-based AutoRec model. We use plate notation to indicate that there are n copies of the neural network (one for each item), where $\mathbf{W}$ and $\mathbf{V}$ are tied across all copies.

$$\min_{\theta} \sum_{\mathbf{r} \in \mathbf{S}} ||\mathbf{r} - h(\mathbf{r}; \theta)||_2^2,$$

$$h(\mathbf{r}; \theta) = f\left(\mathbf{W} \cdot g(\mathbf{V}\mathbf{r} + \boldsymbol{\mu}) + \mathbf{b}\right)$$

$$\min_{\theta} \sum_{i=1}^{n} ||\mathbf{r}^{(i)} - h(\mathbf{r}^{(i)}; \theta))||_{\mathcal{O}}^2 + \frac{\lambda}{2} \cdot (||\mathbf{W}||_F^2 + ||\mathbf{V}||_F^2), \quad (2)$$

$$\hat{R}_{ui} = (h(\mathbf{r}^{(i)}; \hat{\theta}))_u.$$

AutoRec - Evaluation

- Model
  - AutoRec
  - RBM-CF
  - LLORMA
- Dataset
  - Movielens 1M, 10M, Netflix datasets
- Metric: RMSE

AutoRec - Evaluation

	ML-1M	ML-10M
U-RBM	0.881	0.823
I-RBM	0.854	0.825
U-AutoRec	0.874	0.867
I-AutoRec	**0.831**	**0.782**

(a)

$f(\cdot)$	$g(\cdot)$	RMSE
Identity	Identity	0.872
Sigmoid	Identity	0.852
Identity	Sigmoid	**0.831**
Sigmoid	Sigmoid	0.836

(b)

	ML-1M	ML-10M	Netflix
BiasedMF	0.845	0.803	0.844
I-RBM	0.854	0.825	-
U-RBM	0.881	0.823	0.845
LLORMA	0.833	**0.782**	0.834
I-AutoRec	**0.831**	**0.782**	**0.823**

(c)

Table 1: (a) Comparison of the RMSE of I/U-AutoRec and RBM models. (b) RMSE for I-AutoRec with choices of linear and nonlinear activation functions, Movielens 1M dataset. (c) Comparison of I-AutoRec with baselines on MovieLens and Netflix datasets. We remark that I-RBM did not converge after one week of training. LLORMA's performance is taken from [2].

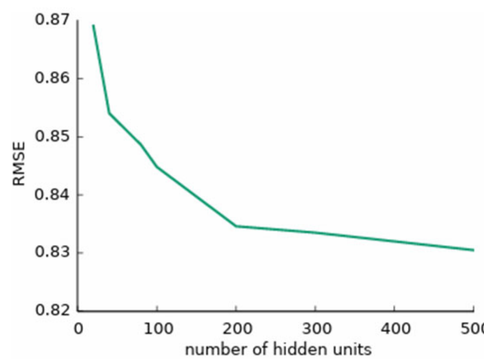


Figure 2: RMSE of I-AutoRec on Movielens 1M as the number of hidden units k varies.

ConvMF - Intro

Convolutional Matrix Factorization for Document Context-Aware Recommendation

Donghyun Kim[1], Chanyoung Park[1], Jinoh Oh[1], Sungyoung Lee[2], Hwanjo Yu[*1]
[1]Pohang University of Science and Technology (POSTECH), Pohang, South Korea
[2]Kyung Hee University, Seoul, South Korea
{[1]kdh5377, [1]pcy1302, [1]kurin, [1]hwanjoyu}@postech.ac.kr, [2]sylee@oslab.khu.ac.kr

ConvMF - Intro

- Rating matrix becomes sparse -> MF becomes inaccurate

  -> Improving accuracy by additionally utilizing textual information

    e.g., reviews, abstracts, synopses

  - LDA, SDAE -> Bag of words models
    - Ignore “contextual information” of document such as surrounding words and word orders.

”people **trust** the man” vs “people betray his **trust** finally”

  -> PMF + CNN

ConvMF - Background

1. Matrix Factorization

$$\mathcal{L} = \sum_{i}^{N}\sum_{j}^{M} I_{ij}(r_{ij} - u_i^T v_j)^2 + \lambda_u \sum_{i}^{N} \|u_i\|^2 + \lambda_v \sum_{j}^{M} \|v_j\|^2$$

2. CNN
   - convolution layer for generating local feature
   - pooling layer for representing data as more concise way

ConvMF - Model

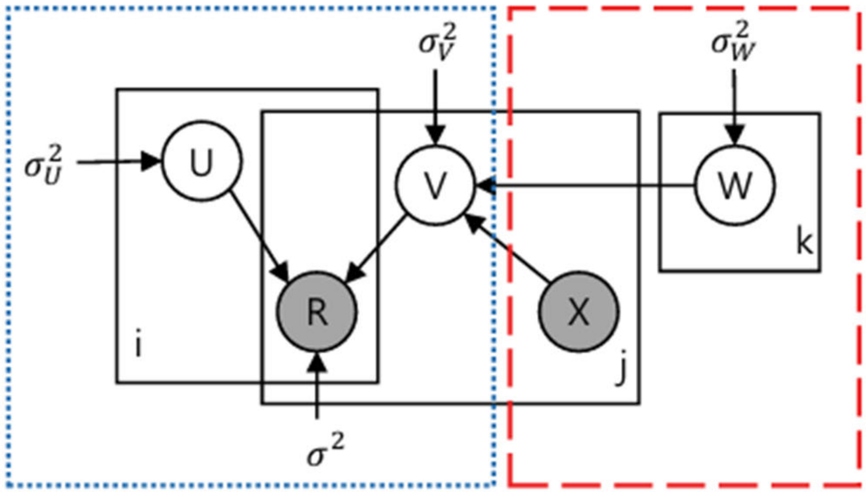


Figure 1: Graphical model of ConvMF model: PMF part in left (dotted-blue); CNN part in right (dashed-red)

ConvMF - Probabilistic Model

$$p(R|U,V,\sigma^2) = \prod_i^N \prod_j^M N(r_{ij}|u_i^T v_j, \sigma^2)^{I_{ij}}$$

$$p(U|\sigma_U^2) = \prod_i^N N(u_i|0, \sigma_U^2 I)$$

$$p(V|W,X,\sigma_V^2) = \prod_j^M N(v_j|cnn(W,X_j), \sigma_V^2 I)$$

$$v_j = cnn(W,X_j) + \epsilon_j$$

$$\epsilon_j \sim N(0, \sigma_V^2 I)$$

$$p(W|\sigma_W^2) = \prod_k N(w_k|0, \sigma_W^2)$$

ConvMF - CNN

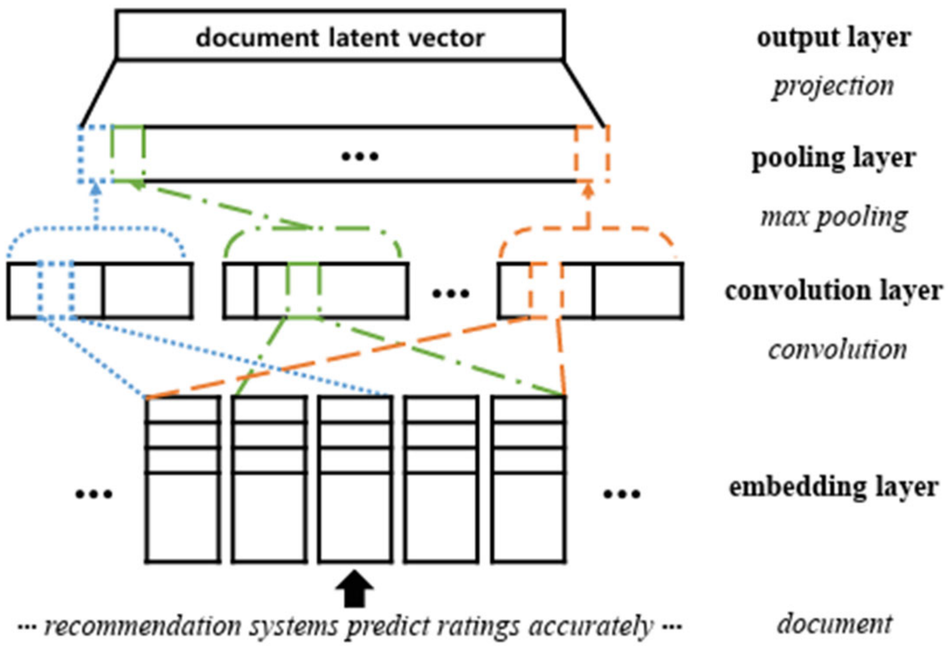


Figure 2: Our CNN architecture for ConvMF

$$s_j = cnn(W, X_j)$$

$$s = \tanh(W_{f_2}\{\tanh(W_{f_1} d_f + b_{f_1})\} + b_{f_2})$$

$$d_f = [\max(c^1), \max(c^2), \ldots, \max(c^j), \ldots, \max(c^{n_c})]$$

$$c^j = [\, c_1^j,\ c_2^j,\ \ldots,\ c_i^j,\ \ldots,\ c_{l-ws+1}^j]$$
$$c_i^j = f(W_c^j * D_{(:,i:(i+ws-1))} + b_c^j)$$

$$D = \begin{bmatrix} \cdots & | & | & | & \cdots \\ & w_{i-1} & w_i & w_{i+1} & \\ & | & | & | & \end{bmatrix}$$

ConvMF - Optimization

- MAP

$$\max_{U,V,W} p(U,V,W|R,X,\sigma^2,\sigma_U^2,\sigma_V^2,\sigma_W^2)$$
$$= \max_{U,V,W} [p(R|U,V,\sigma^2)p(U|\sigma_U^2)p(V|W,X,\sigma_V^2)p(W|\sigma_W^2)]$$

$$\mathcal{L}(U,V,W) = \sum_i^N \sum_j^M \frac{I_{ij}}{2}(r_{ij} - u_i^T v_j)_2 + \frac{\lambda_U}{2}\sum_i^N \|u_i\|_2$$
$$+ \frac{\lambda_V}{2}\sum_j^M \|v_j - cnn(W,X_j)\|_2 + \frac{\lambda_W}{2}\sum_k^{|w_k|}\|w_k\|_2,$$

  - Coordinate descent per iteration

$$u_i \leftarrow (VI_iV^T + \lambda_U I_K)^{-1} VR_i$$
$$v_j \leftarrow (UI_jU^T + \lambda_V I_K)^{-1}(UR_j + \lambda_V cnn(W,X_j))$$

ConvMF - Optimization

$$\mathcal{E}(W) = \frac{\lambda_V}{2} \sum_{j}^{M} \|(v_j - cnn(W, X_j))\|^2 + \frac{\lambda_W}{2} \sum_{k}^{|w_k|} \|w_k\|^2 + \text{constant}$$

Optimize W : Backpropagation with given target value v_j

$$\begin{aligned} r_{ij} &\approx \mathbb{E}[r_{ij}|u_i^T v_j, \sigma^2] \\ &= u_i^T v_j = u_i^T (cnn(W, X_j) + \epsilon_j) \end{aligned}$$

ConvMF - Experiment

- Model
  - PMF
  - CTR
  - CDL
  - ConvMF
  - ConvMF+

Dataset	# users	# items	# ratings	density
ML-1m	6,040	3,544	993,482	4.641%
ML-10m	69,878	10,073	9,945,875	1.413%
AIV	29,757	15,149	135,188	0.030%

Table 1: Data statistic on three real-world datasets

- Dataset
  - MovieLens 1m, 10m, Amazon Instant Video(AIV)

- Metric: RMSE

ConvMF - Experiment

Model	Dataset		
	ML-1m	ML-10m	AIV
PMF	0.8971	0.8311	1.4118
CTR	0.8969	0.8275	1.5496
CDL	0.8879	0.8186	1.3594
ConvMF	**0.8531**	**0.7958**	**1.1337**
ConvMF+	**0.8549**	**0.7930**	**1.1279**
Improve	3.92%	2.79%	16.60%

Table 3: Overall test RMSE

ConvMF - Experiment

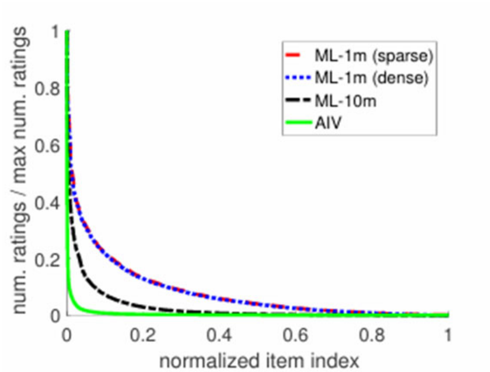


Figure 3: Skewness of the number of ratings for items on each dataset

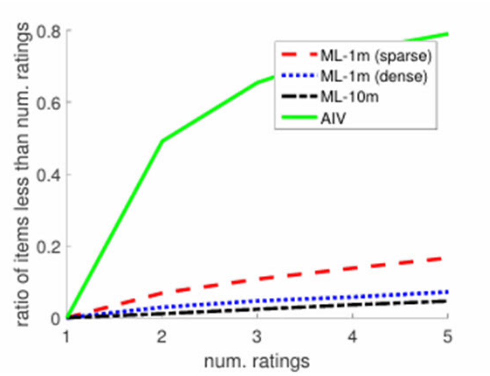


Figure 4: Ratio of items that have less than num. ratings (N) to each entire dataset

Model	ML-1m		ML-10m		AIV	
	λ_U	λ_V	λ_U	λ_V	λ_U	λ_V
PMF	0.01	10000	10	100	0.1	0.1
CTR	100	1	10	100	10	0.1
CDL	10	100	100	10	0.1	100
ConvMF	100	10	10	100	1	100
ConvMF+	100	10	10	100	1	100

Table 2: Parameter Setting of λ_U and λ_V

ConvMF - Experiment

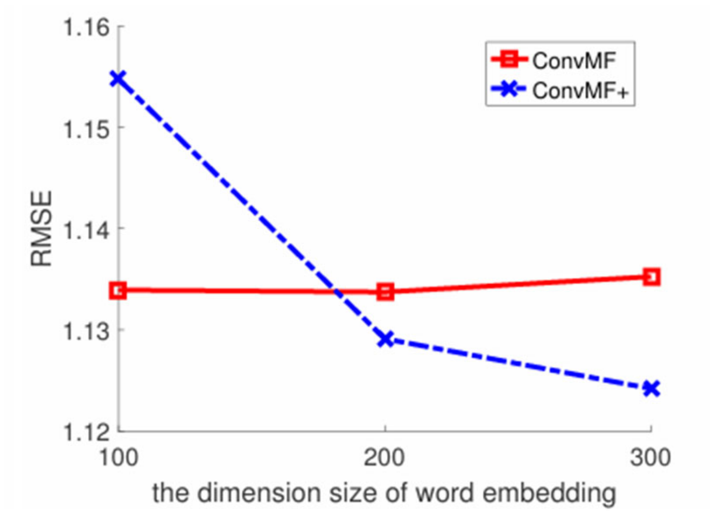


Figure 5: The effects of the dimension size of word embedding on Amazon dataset

ConvMF - Experiment

Verb

noun

Phrase captured by W_c^{11}	$\max(c^{11})$	Phrase captured by W_c^{86}	$\max(c^{86})$
people **trust** the man	***0.0704***	betray his **trust** finally	***0.1009***
Test phrases for W_c^{11}	c_{test}^{11}	Test phrases for W_c^{86}	c_{test}^{86}
people **believe** the man	***0.0391***	betray his **believe** finally	*0.0682*
people **faith** the man	*0.0374*	betray his **faith** finally	***0.0693***
people **tomas** the man	*0.0054*	betray his **tomas** finally	*0.0480*

Table 5: Case study on two shared weights of ConvMF

Conclusion

- AutoRec
  - CF + NN(AutoEncoder)
- ConvMF
  - PMF + CNN