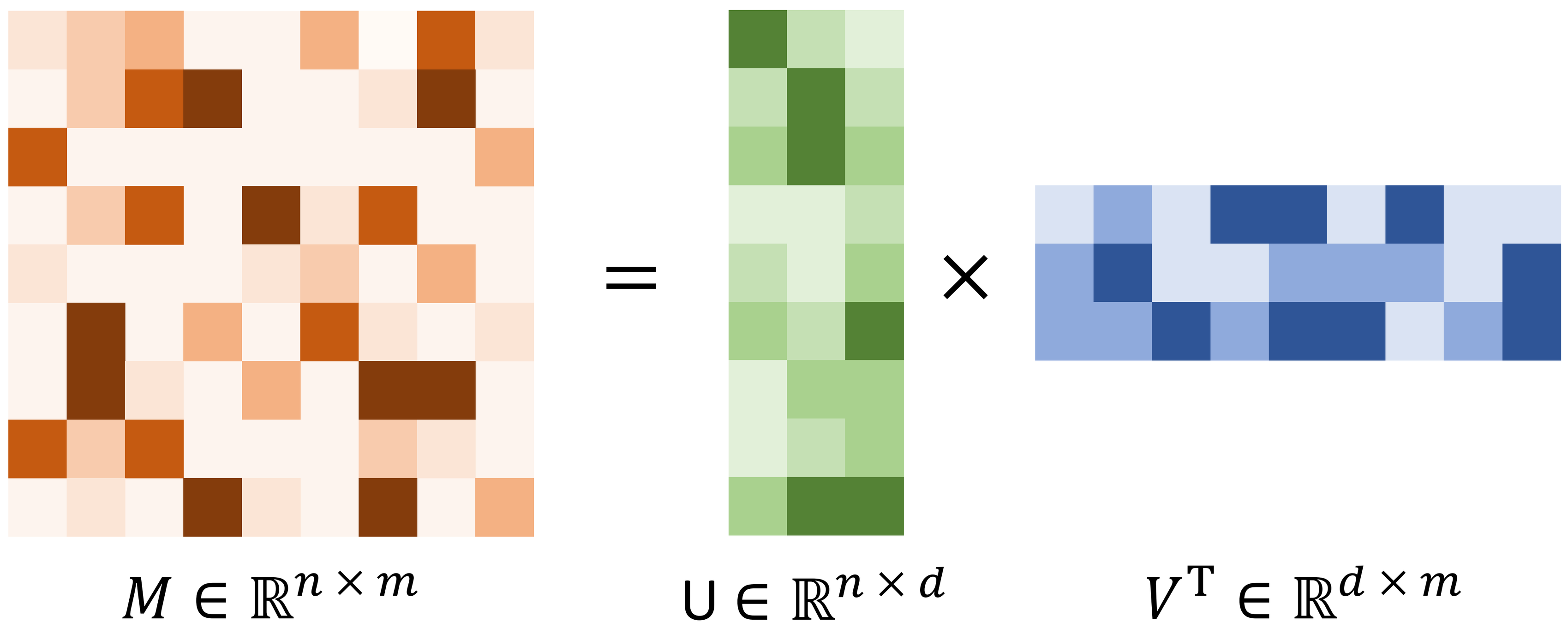
EXTENDABLE NEURAL MATRIX COMPLETION

Duc Minh Nguyen, Evaggelia Tsiligianni, Nikos Deligiannis
Vrije Universiteit Brussel – imec
{mdnguyen,etsiligi,ndeligia}@etrovub.be

VUB VRIJE UNIVERSITEIT BRUSSEL imec

1. Problem Statement



- **Goal**: Matrix Completion ***by learning factorizations***
- Learn factors $U_i, i = 1 \dots n$, and $V_j, j = 1 \dots m$, by minimizing:

$$L = \left\| P_\Omega(UV^T) - P_\Omega(M) \right\|_F^2$$

Ω: the set containing indices of known entries in M

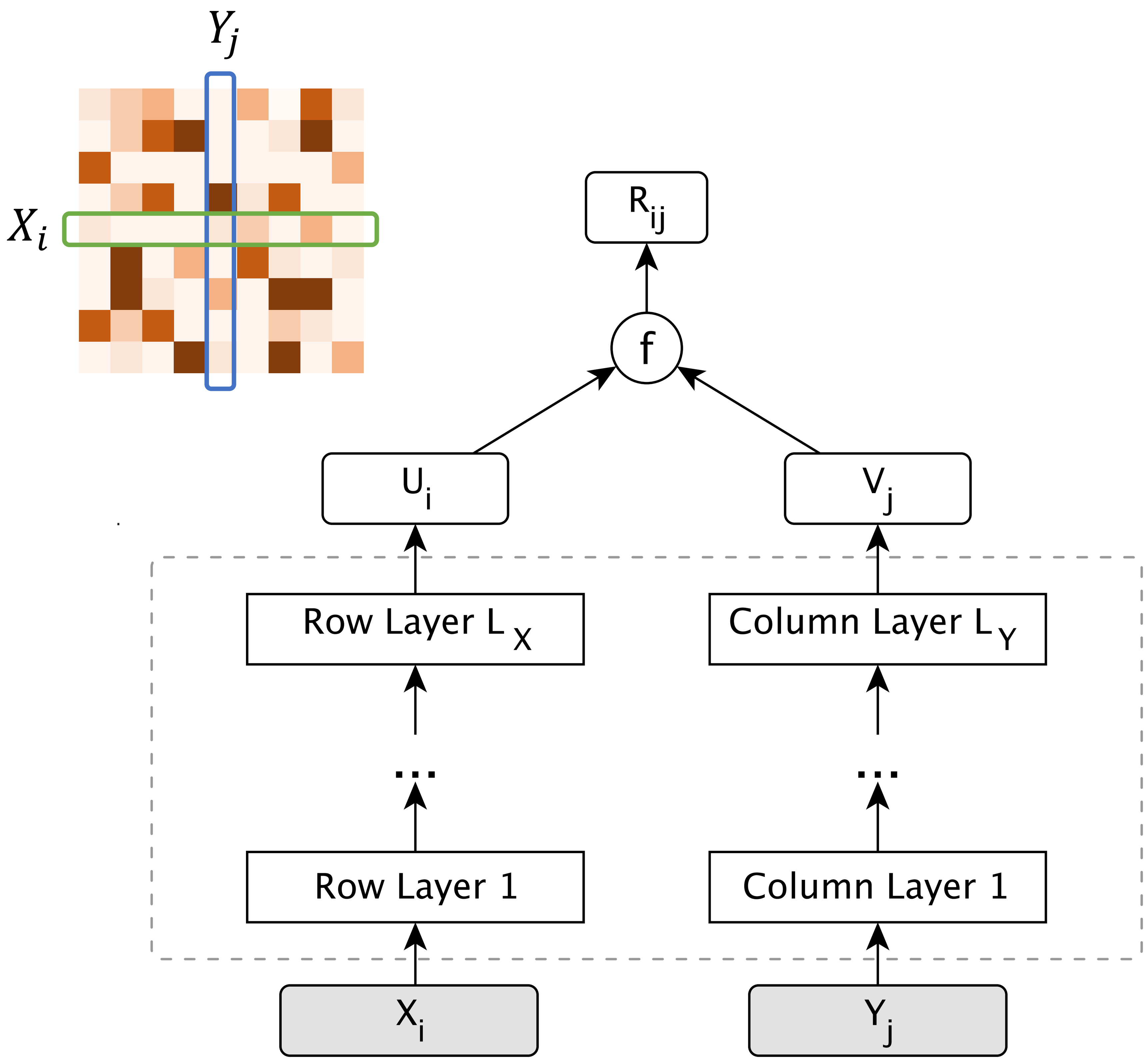
P_Ω: an operator that indexes the entries defined by Ω

- State-of-the-art results in Matrix Completion are achieved using ***neural networks***
- **Problem**: Existing models ***do not extend well to unseen rows and columns***

*Can we build a model **extendable** to **unseen** rows & columns (can calculate U_t, V_k with $t > n, k > m$ **without retraining**)?*

2. Proposed Model

Neural Matrix Completion (NMC)

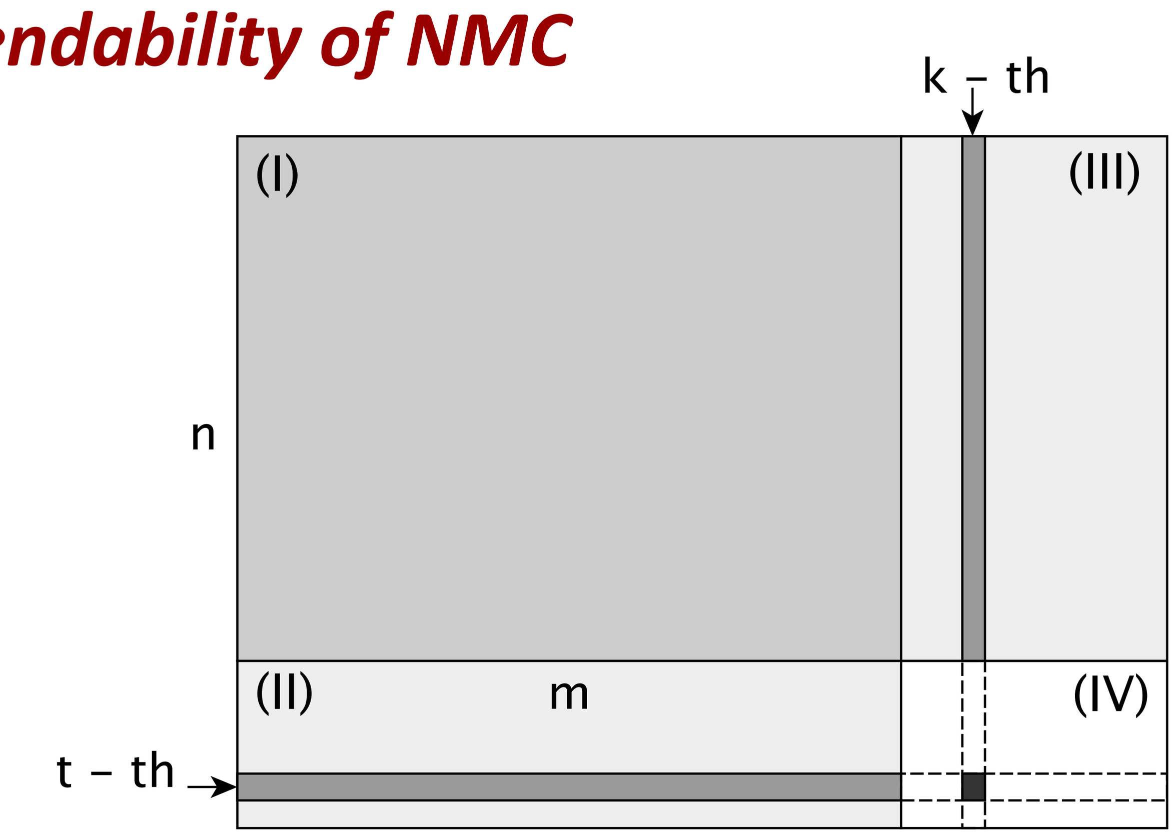


- R_{ij} : prediction for matrix entry M_{ij}, f is cosine similarity

$$f(U_i, V_j) = \frac{U_i^T V_j}{\|U_i\|_2 \|V_j\|_2}$$

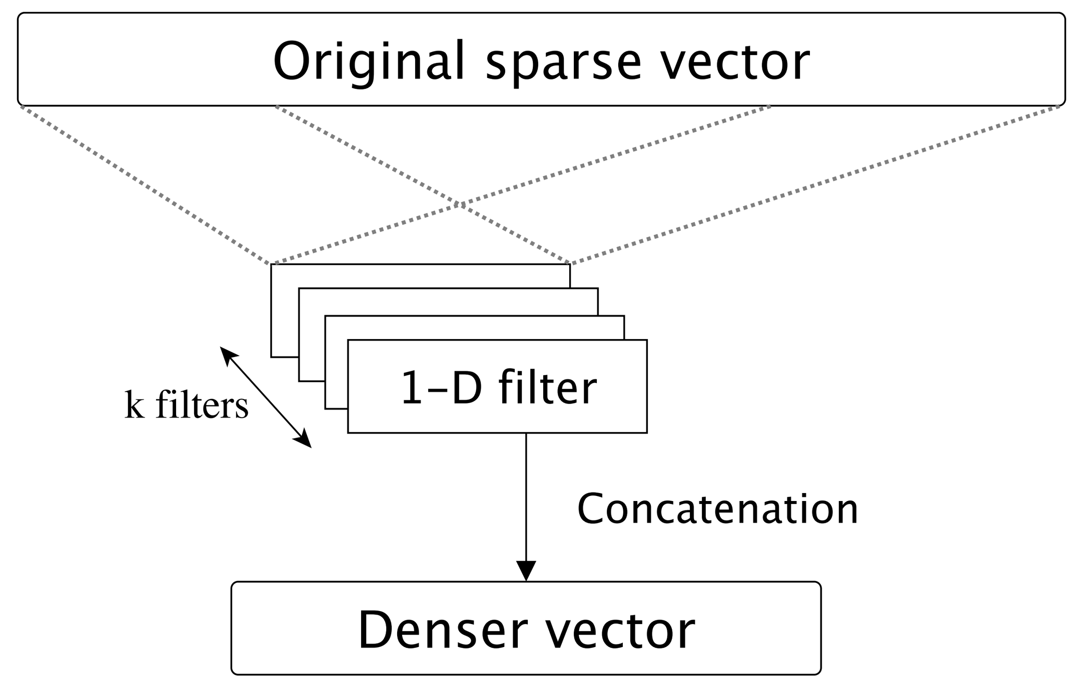
- Each network branch consists of ***fully connected layers***
- Learn U_i, V_j from ***row and column vectors*** (X_i, Y_j)
- Inherits the ***expressive power*** of deep networks for learning representations

Extendability of NMC



- Area (I): ***incomplete*** rows and columns available during training
- Areas (II), (III), (IV): ***new rows and columns*** (from which only ***some entries*** are ***observed after training***)
- Matrices $M^{(I)}, M^{(I)\&(II)}, M^{(I)\&(III)}, M^{(I)-(IV)}$ correspond to area (I), areas (I) & (II), areas (I) & (III) and areas (I), (II), (III), (IV) respectively.
- With $i = 1 \dots n, j = 1 \dots m$, $\boldsymbol{t > n, k > m}$, NMC predicts:
  - ${M^{(I)\&(II)}}_{tj}$ by taking as inputs $X_t = {M^{(I)\&(II)}}_{t,:}$, $Y_j = {M^{(I)}}_{:,j}$
  - ${M^{(I)\&(III)}}_{ik}$ by taking as inputs $X_i = {M^{(I)}}_{i,:}$, $Y_k = {M^{(I)\&(III)}}_{:,k}$
  - ${M^{(I)-(IV)}}_{tk}$ by taking as inputs $X_t = {M^{(I)\&(II)}}_{t,:}$, $Y_k = {M^{(I)\&(III)}}_{:,k}$

Scaling NMC to Large-Scale data



- Fully connected layers are not efficient for ***high dimensional***, and ***highly sparse*** matrices (e.g. Netflix movie ratings)
- ***Solution***: Add ***summarization layers*** (1-D convolutional layers) before fully connected layers

3. Experimental Results

Table 1. Results on the MovieLens1M dataset

	Area (I)		Area (II)		Area (III)		Area (IV)	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
U-CF-NADE-S [1]	0.855	0.671	-	-	-	-	-	-
I-CF-NADE-S [1]	**0.839**	**0.651**	-	-	-	-	-	-
U-Autorec [2]	0.906	0.772	0.976	0.781	-	-	-	-
I-Autorec [2]	0.841	0.662	-	-	**0.856**	**0.670**	-	-
Deep U-Autorec [3]	0.889	0.702	0.969	0.765	-	-	-	-
Proposed	0.850	0.675	**0.883**	**0.699**	0.864	0.685	**0.904**	**0.715**

Table 2. Results on the Netflix dataset

	Area (I)		Area (II)		Area (III)		Area (IV)	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
I-Autorec [2]	0.842	0.655	-	-	**0.862**	**0.671**	-	-
Deep U-Autorec [3]	0.848	0.662	0.879	0.689	-	-	-	-
Proposed	0.856	0.676	**0.861**	**0.680**	0.873	0.688	**0.877**	**0.692**

References

[1] Zheng et al., "A neural autoregressive approach to collaborative filtering", ICML 2016
[2] Sedhain et al., "Autorec: Autoencoders meet collaborative filtering", WWW 2015
[3] Kuchaiev et al., "Training deep autoencoders for collaborative filtering", arXiv 2017