Probabilistic Matrix Factorization with Non-random Missing Data

José Miguel Hernández–Lobato[1],

June 23, 2014

joint work with

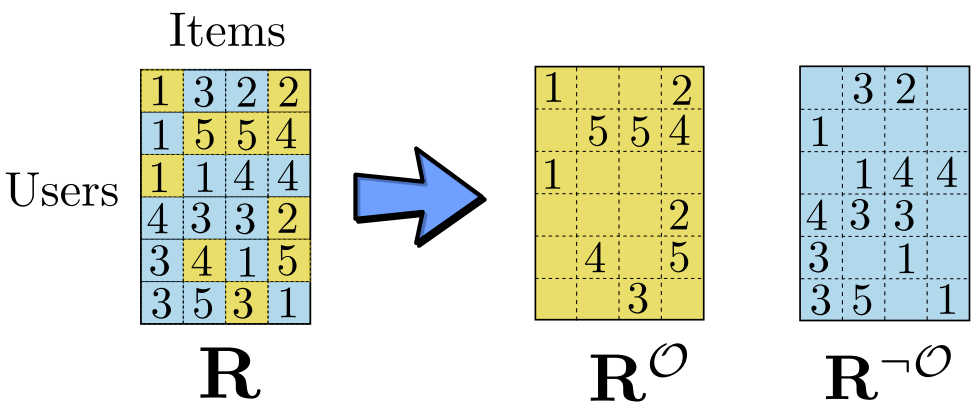
Neil Houlsby[1] and Zoubin Ghahramani[1]

[1]Department of Engineering, Cambridge University.

Rating Data and Matrices

Large amounts of **rating data** produced by media streaming, news, retailing, urban city guides, scientific conferences, etc...

The ratings are encoded as a matrix $\mathbf{R}$. In practice, few matrix entries are observed ($\mathbf{R}^{\mathcal{O}}$) and **many are missing** ($\mathbf{R}^{\neg\mathcal{O}}$).



Matrix Factorization Models

Main assumption: $\mathbf{R}$ can be well approximated by the **low rank** matrix $\mathbf{U}\mathbf{V}^{\mathrm{T}}$, $\mathbf{U}$ and $\mathbf{V}$ have a small number of columns.

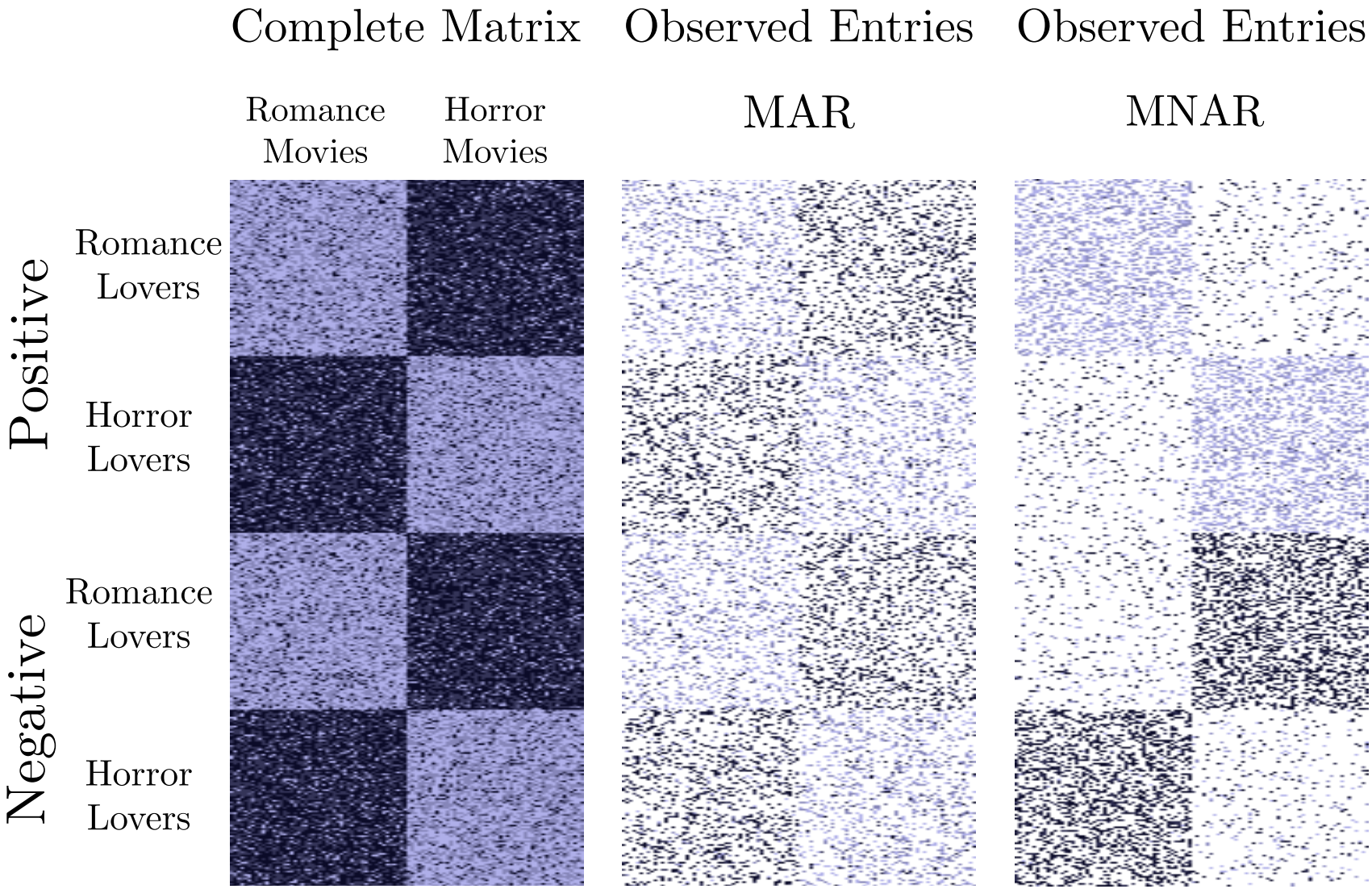
$$\mathbf{R} = \begin{bmatrix} 1 & 3 & 2 & 2 \\ 1 & 5 & 5 & 4 \\ 1 & 1 & 4 & 4 \\ 4 & 3 & 3 & 2 \\ 3 & 4 & 1 & 5 \\ 3 & 5 & 3 & 1 \end{bmatrix} \approx \mathbf{U} \; \mathbf{V}^{\mathrm{T}}$$

We can **marginalize out** the missing entires in the likelihood:

$$p(\mathbf{R}^{\mathcal{O}}|\mathbf{U},\mathbf{V}) = \sum_{\mathbf{R}^{\neg\mathcal{O}}} \left[\prod_{i=1}^{n} \prod_{j=1}^{d} p(r_{i,j}|\mathbf{u}_i\mathbf{v}_j^{\mathrm{T}}) \right] = \prod_{(i,j)\in\mathcal{O}} p(r_{i,j}|\mathbf{u}_i\mathbf{v}_j^{\mathrm{T}}) \,.$$

Introduces independencies and MAR assumption!
The data may be MNAR!

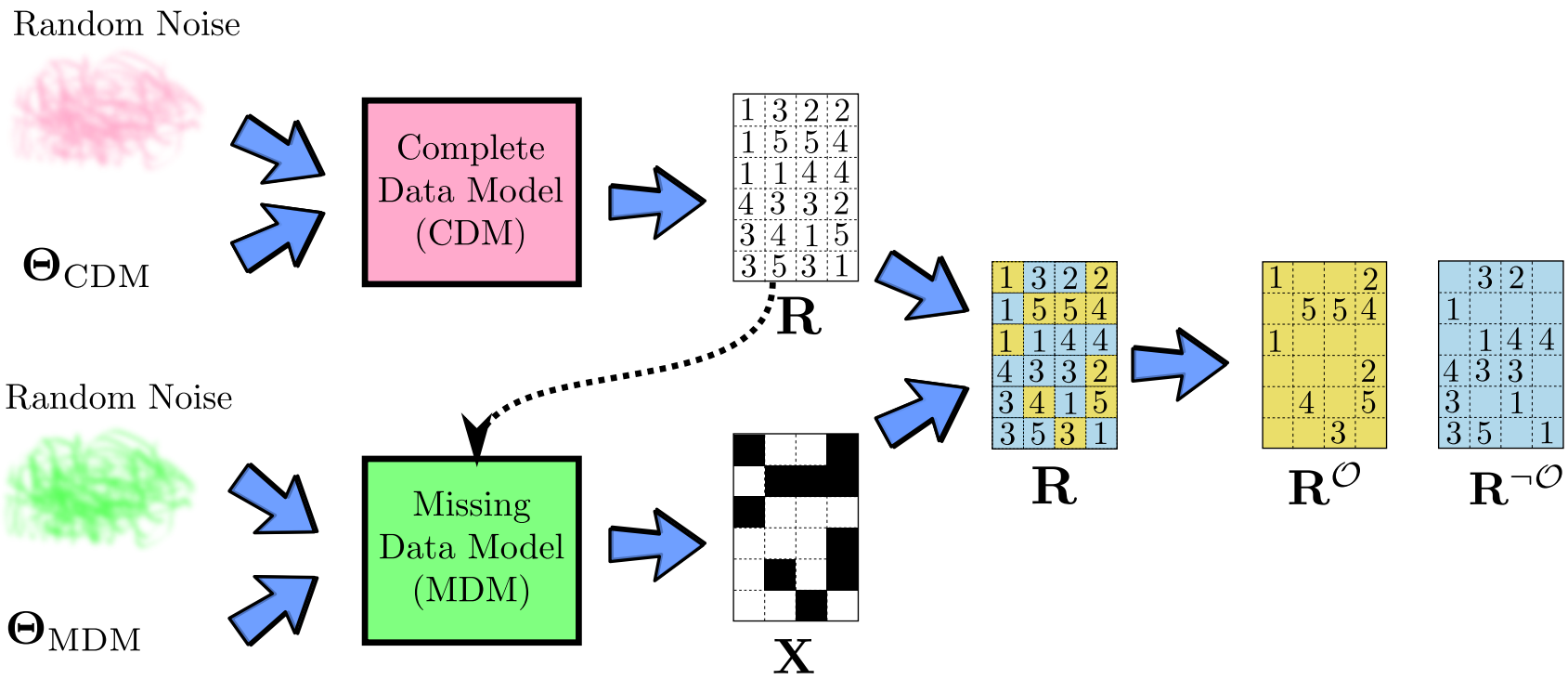
The Synthetic SRH Dataset



Most matrix factorization models assume MAR data.

What to do if data is MNAR?

Probabilistic Treatment of the MNAR Data



The joint likelihood for $\boldsymbol{\Theta}_{\text{CDM}}$ and $\boldsymbol{\Theta}_{\text{MDM}}$ given $\mathbf{X}$ and $\mathbf{R}$ is

$$p(\mathbf{X}, \mathbf{R}|\boldsymbol{\Theta}_{\text{CDM}}, \boldsymbol{\Theta}_{\text{MDM}}) = p(\mathbf{X}|\mathbf{R}, \boldsymbol{\Theta}_{\text{MDM}})p(\mathbf{R}|\boldsymbol{\Theta}_{\text{CDM}}) .$$

Missing Data Model

Complete Data Model

Missing Data Model (MDM)

A state-of-the-art matrix factorization model:

- Logistic likelihood, global bias and latent factors.
- The **effect** of $r_{i,j}$ on $x_{i,j}$ changes across rows and columns.
- **SVI** for scalable inference.

The likelihood function is

$$p(\mathbf{X}|\mathbf{E},\mathbf{F},z,\mathbf{\Lambda}^{\text{row}},\mathbf{\Psi}^{\text{col}},\mathbf{R}) = \left[\prod_{i=1}^{n}\prod_{j=1}^{d}\sigma\{(2x_{i,j}-1)(\mathbf{e}_i\mathbf{f}_j^{\text{T}} + z + \sum_{l=1}^{L}(\lambda_{i,l}^{\text{row}} + \psi_{j,l}^{\text{col}})\mathbf{I}[r_{i,j}=l])\}\right].$$

user effect of $r_{i,j}$

item effect of $r_{i,j}$

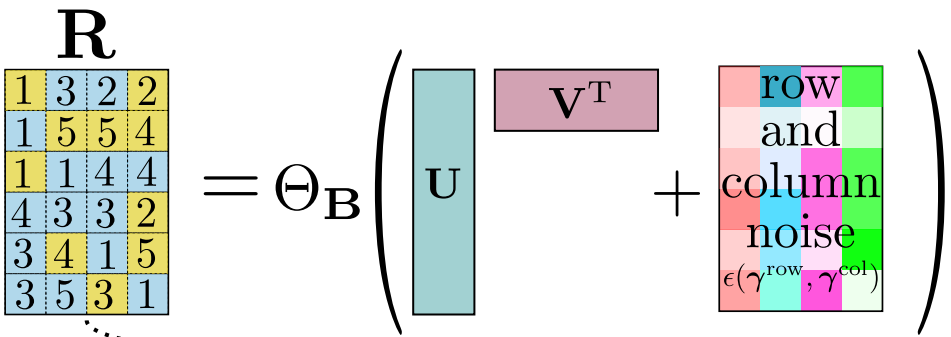
logistic function

latent factors

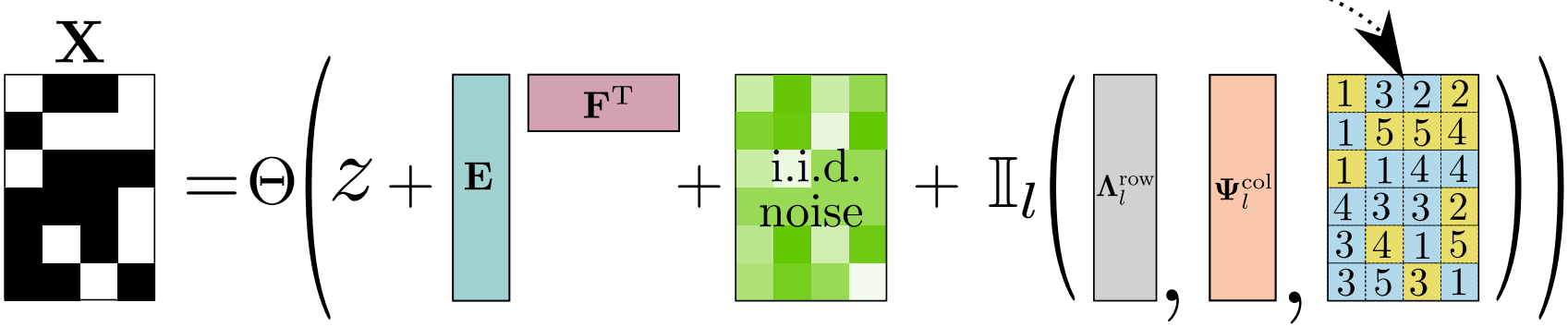
global bias

A Sketch of the Joint Model

Complete Data Model (CDM):



Missing Data Model (MDM):



Inference Algorithm and Predictive Distribution

We adjust the **posterior approximation** $\mathcal{Q}$ using

Input: Rating dataset $\mathcal{D}$.

Individual Learning of CDM and MDM.

Adjust $\mathcal{Q}$ using EP and VB on CDM under MAR assumption.
Adjust $\mathcal{Q}$ using SVI on MDM ignoring CDM.

for $t = 1$ **to** T **do**

Joint Learning of CDM and MDM.

Adjust $\mathcal{Q}$ using SVI on MDM with CDM predictions.
Adjust $\mathcal{Q}$ using EP-SVI on CDM with MDM predictions.

end for

Output: Posterior approximation $\mathcal{Q}$.

Given $\mathcal{Q}$, the **predictive distribution** for $r_{i,j}$ depends on $x_{i,j}$:

$$p(r_{i,j} = l|\mathbf{R}^{\mathcal{O}}, \mathbf{X}) \approx \tilde{p}^{\mathrm{JM}}_{i,j,l}(x_{i,j}) \propto \tilde{p}^{\mathrm{CDM}}_{i,j,l}\tilde{p}^{\mathrm{MDM}}_{i,j,l}(x_{i,j}),$$

Predictive Performance on the Rating Matrix $\mathbf{R}$

Test sets: 1% of ratings $r_{i,j}$with $x_{i,j} = 1$.
$x_{i,j}$ is set to **zero** for these entries.

Table : Average Log-likelihood on the Test Sets.

	Dataset	MF MNAR	MF MAR	MM MAR	CTPv MNAR	Logitvd MNAR	Paquet MAR	Oracle
Real-world	ML100K	**-1.181**	-1.186	-1.471	-1.463	-1.425	-1.218	-1.468
	ML1M	**-1.121**	-1.125	-1.308	-1.436	-1.380	-1.162	-1.456
	MTweet	**-0.941**	-0.946	-1.105	-1.245	-1.141	-0.997	-1.235
	NIPS	**-0.937**	-0.956	-1.204	-1.170	-1.167	-0.995	-1.329
	Yahoo	**-1.172**	-1.204	-1.278	-1.399	-1.304	-1.218	-1.551
Synthetic	SMF-MNAR	**-0.902**	-0.937	-1.447	-1.336	-1.326	-1.000	-1.331
	SMF-MAR	-0.425	**-0.417**	-1.327	-1.238	-1.235	-0.510	-1.198
	SRH-MNAR	-1.055	-1.067	-0.987	**-0.962**	-0.963	-1.143	-1.392
	SRH-MAR	-1.317	-1.287	-1.272	**-1.265**	-1.266	-1.318	-1.498

MF-MNAR: CDM & MDM, MNAR assumption.
MF-MAR: CDM only, MAR assumption.

Predictive Performance on the Binary Matrix $\mathbf{X}$

Test sets: 1% of ratings $r_{i,j}$ with $x_{i,j} = 1$.

$x_{i,j}$ is set to **zero** for these entries.

Objective: find the $x_{i,j}$ that took value 1 and were flipped.

Table : Average Recall.

	Dataset	MF MNAR	MDM	CTPv MNAR	Logitvd MNAR	Freq
Real-world	ML100K	**0.299**	0.295	0.093	0.130	0.119
	ML1M	0.204	**0.204**	0.041	0.068	0.077
	MTweet	**0.203**	0.199	0.127	0.142	0.143
	NIPS	**0.309**	0.309	0.011	0.009	0.013
	Yahoo	**0.285**	0.283	0.145	0.198	0.182
Synthetic	SMF-MNAR	**0.300**	0.280	0.038	0.052	0.051
	SMF-MAR	0.438	**0.445**	0.038	0.051	0.047
	SRH-MNAR	**0.246**	0.245	0.209	0.157	0.121
	SRH-MAR	0.113	0.115	0.134	**0.143**	0.131

MF-MNAR: CDM & MDM.

MDM: MDM only.

Conclusions

- We have proposed a probabilistic matrix factorization model for the MNAR setting.
- Our MDM is the first one that uses a matrix factorization model to account for patterns in $\mathbf{X}$ not related to $\mathbf{R}$.
- In the proposed MDM, the effect of $r_{i,j}$ on $x_{i,j}$ changes across rows and across columns.
- The combination of the proposed MDM and CDM yields gains in the predictions for both $\mathbf{X}$ and $\mathbf{R}$.
- We achieve scalability using stochastic inference methods.

Thank you for your attention!

Average Accuracy on **R**

Table : Average Accuracy on the Test Sets.

Dataset	MF MNAR	MF MAR	MM MAR	CTPv MNAR	Logitvd MNAR	Paquet MAR	Oracle
ML100K	**0.479**	0.476	0.346	0.370	0.374	0.465	0.335
ML1M	**0.499**	0.495	0.416	0.397	0.400	0.486	0.348
MTweet	0.603	**0.603**	0.547	0.515	0.530	0.589	0.492
NIPS	**0.612**	0.609	0.455	0.495	0.494	0.594	0.398
Yahoo	**0.536**	0.522	0.493	0.437	0.481	0.503	0.314
SMF-MNAR	**0.638**	0.632	0.438	0.467	0.464	0.554	0.448
SMF-MAR	0.829	**0.835**	0.424	0.439	0.437	0.819	0.464
SRH-MNAR	0.477	0.476	**0.492**	0.474	0.480	0.475	0.318
SRH-MAR	0.419	0.430	0.422	0.419	0.416	**0.446**	0.293

Average MAE on **R**

Table : Average MAE on the Test Sets.

Dataset	MF MNAR	MF MAR	MM MAR	CTPv MNAR	Logitvd MNAR	Paquet MAR	Oracle
ML100K	**0.639**	0.642	0.826	0.806	0.801	0.645	0.900
ML1M	**0.595**	0.598	0.716	0.768	0.752	0.598	0.872
MTweet	0.451	**0.451**	0.534	0.553	0.540	0.457	0.593
NIPS	**0.532**	0.544	0.776	0.713	0.707	0.548	1.011
Yahoo	0.814	0.825	0.903	1.126	0.944	**0.800**	1.401
SMF-MNAR	**0.435**	0.459	0.832	0.824	0.819	0.466	0.935
SMF-MAR	0.170	**0.165**	0.675	0.623	0.620	0.183	0.602
SRH-MNAR	0.713	0.701	0.717	0.689	**0.684**	0.698	1.447
SRH-MAR	0.808	0.793	0.818	0.819	0.823	**0.782**	1.401