On incorporating auxiliary information into recommender systems using adversarial training

Dora Jambor • Putra Manggala

Shopify Inc., Montreal, Canada

Key Question

How to incorporate auxiliary information into recommender systems such that ...

- domain expertise from multiple sources and diversity can be induced into the recommendations,
- recommendation accuracy is high,
- recommendation model is interpretable?

Linear Recommenders

Linear methods like SLIM [1] and LRec [2] produce accurate and interpretable recommendations by learning similarity metric from data.

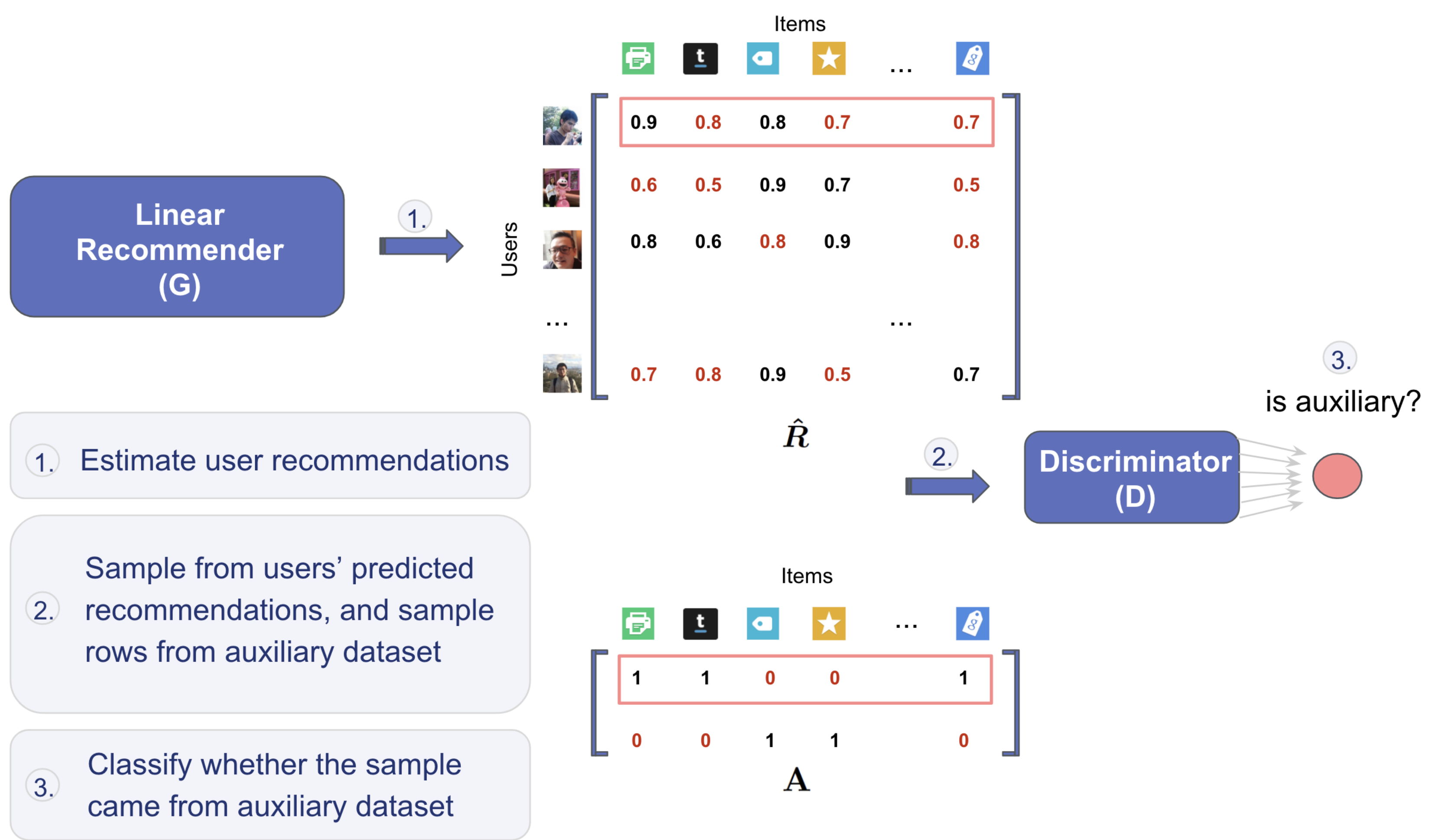
$$\min_{\mathbf{W} \in C} \|\mathbf{R} - \mathbf{RW}\|_F^2 + \frac{\lambda}{2}\|\mathbf{W}\|_F^2 + \mu\|\mathbf{W}\|_1$$

$$C = \{\mathbf{W} \in \mathbb{R}^{nxn} : diag(\mathbf{W}) = 0, \mathbf{W} \geq 0\}$$

The AdRec Model

Auxiliary information can be encoded as an indicator vector of items:

Using the adversarial framework from [3], we iteratively optimize: G, the linear recommender and D, the discriminator.



$$\min_G \max_D \mathbb{E}_{\mathbf{x} \sim p_{aux}(\mathbf{x})}[log D(\mathbf{x})] + \mathbb{E}_{\mathbf{r} \sim p_{data}(\mathbf{r})}[log(1 - D(G(\mathbf{r})))] \quad (1)$$

$$+\mathbb{E}_{\mathbf{r} \sim p_{data}(\mathbf{r})}[\|\mathbf{r} - G(\mathbf{r})\|_2 + \Omega(\mathbf{W}_G)], \quad (2)$$

Data

For our experiments, R training matrix is built using 3000 users' implicit item interactions with 3437 Shopify applications. To reduce sparsity, we omit users with less than 4 interactions. An auxiliary dataset A is composed of three indicator vectors which represent auxiliary information.

Preliminary results

Experiments were conducted to demonstrate a useful trade-off between two metrics, precision (p@k) and auxiliary precision (p_{aux}@k), defined as:

$$\text{p@k} = \frac{|\{relevant\ items\} \bigcap \{recommended\ items[:k]\}|}{k}$$

$$\text{p}_{aux}\text{@k} = \frac{|\{auxiliary\ items\} \bigcap \{recommended\ items[:k]\}|}{k}$$

A few important hyperparameters for the trade-off:

- γ, a learning rate for G's adversarial update (2nd term in (1)),
- α, one-sided label noise,
- number of epochs

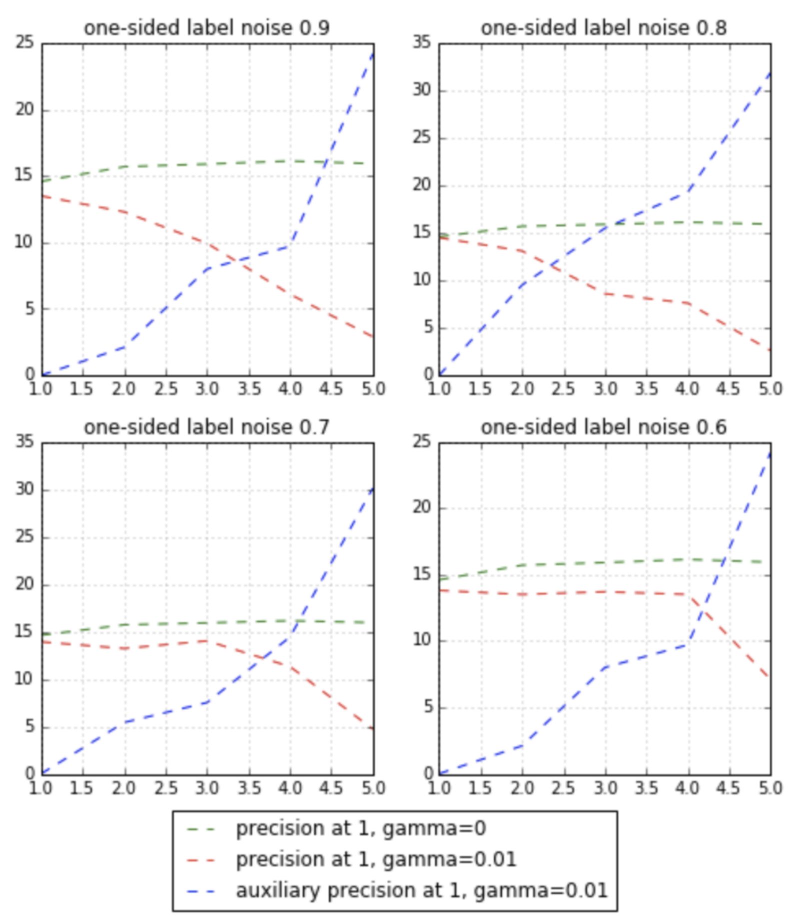
Table 1. Trade-off behavior for γ and α given 3 epochs.

γ	α	p@1	p_{aux}@1	p@3	p_{aux}@3
0.0	-	15.5%	0	9%	0
0.01	0.6	14.3%	7.8%	8.4%	11.0%
0.01	0.7	12.9%	8.0%	7.0%	13.7%
0.01	0.8	10%	15.2%	6.4%	18.9%
0.01	0.9	9.6%	15.1%	6%	19.4%

AdRec incorporates auxiliary information in top-k recommendations

As α increases ↑
- p@k decreases ↓
- p_{aux}@k increases ↑

Figure 1. Precision trade-off for linear and adversarial recommender.



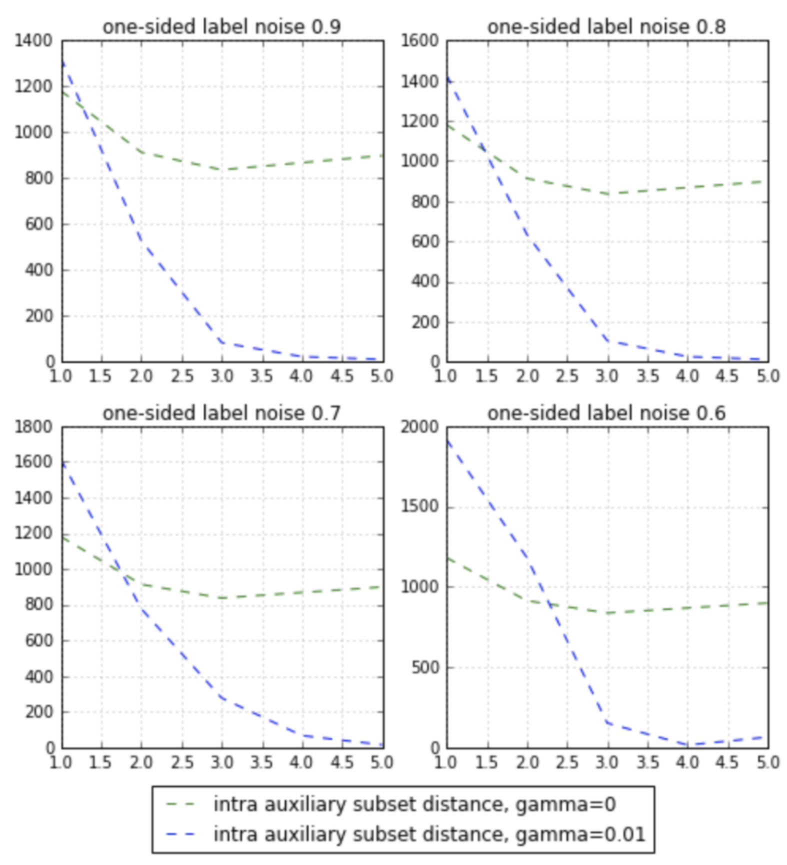
Trade-off between p@1 and p_{aux}@1 increases with the number of epochs.

Green line shows the precision of a linear recommender.

Search the hyperparameter space to find the optimal trade-off profile.

Based on the above, a practitioner would choose the profiles with label noise 0.6 and 0.7.

Figure 2. Intra auxiliary subset distance for linear and adversarial recommender



Intra auxiliary subset distance

Compute recommendations for all users

Consider where the auxiliary items are in the ranked list, then calculate their rank distance

As epochs ↑, for any pair of auxiliary items:

Mean rank distance across all users ↓

Useful debugging metric for AdRec training for user recommendations beyond top-k.

Acknowledgements

We thank Ga Wu for his initial ideas on adversarial recommenders, and his continued collaboration and support.

Bibliography

[1] Ning, Xia, and George Karypis. "Slim: Sparse linear methods for top-n recommender systems." ICDM, 2011.

[2] Sedhain, Suvash, et al. "On the Effectiveness of Linear Models for One-Class Collaborative Filtering." AAAI. 2016.

[3] Goodfellow, Ian, et al. "Generative adversarial nets." NIPS. 2014.