Top-N Recommendation with Missing Implicit Feedback

Daryl Lim
University of California, San Diego

Julian McAuley
University of California, San Diego

Gert Lanckriet
University of California, San Diego

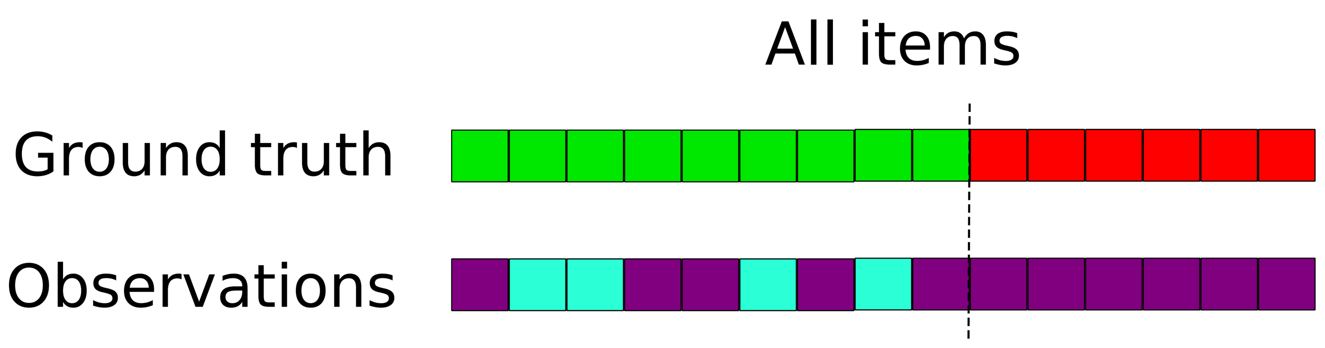
Introduction

Main Contribution

We discuss a missing data model for implicit feedback and propose a novel evaluation measure which is unbiased with respect to the missing data. We also present an efficient algorithm to optimize our measure.

Data Model

For each user, assume that observed relevant items are a simple random sample from some unknown ground truth prior relevant set:



- $\mathcal{P}_u^+$, prior relevant set
- $\mathcal{P}_u^-$, prior irrelevant set
- $\mathcal{O}_u^+$, observed relevant set (drawn from $\mathcal{P}_u^+$)
- $\mathcal{O}_u^-$, observed irrelevant set

To evaluate a given predicted ranking, we would like to have an evaluation measure that, when evaluated on the observed relevant/irrelevant set, returns the same value **in expectation** over all possible patterns of observations.as when evaluated on the prior relevant/irrelevant (ground truth) set, which we call unbiased-to-missing-data (UBM).

In previous work, Steck et. al. have proposed the ATOP and Recall@N measures which meet this criteria. However, ATOP is very similar to AUC while Recall@N is hard to optimize directly.

We want a measure which **is UBM,** focuses on the **top of the ranking** and **can be optimized in an efficient manner.**

Proposed method

Average Discounted Gain (ADG)

We propose the **Average Discounted Gain** measure which has the UBM property.

Let $f_\theta(u,i)$ be the prediction function for a (user, item) pair, and $\mathcal{I}$ be the set of all items.

Let $\mathrm{rank}(i) = \sum_{i' \in \mathcal{I} \setminus i} \mathbf{I}(f_\theta(u,i') - f_\theta(u,i))$

Average Discounted Gain (ADG):

$$\frac{1}{|\mathcal{R}_u^+|} \sum_{i^+ \in \mathcal{R}_u^+} \frac{1}{\log_2(\mathrm{rank}(i^+) + 2)}$$

Optimization

In order to maximize ADG performance, we opt to minimize 1-ADG on the training set.

$$1 - \mathrm{ADG} = 1 - \frac{1}{|\mathcal{O}_u^+|} \sum_{i^+ \in \mathcal{O}_u^+} \frac{1}{\log_2(\mathrm{rank}(i^+) + 2)}$$

$$= \frac{1}{|\mathcal{O}_u^+|} \sum_{i^+ \in \mathcal{O}_u^+} \mathcal{C}(\mathrm{rank}(i^+))$$

where

$$\mathcal{C}(k) = 1 - \frac{1}{\log_2(k+2)}$$

Algorithm 1 The OPT-ADG algorithm

Require: user set $\mathcal{U}$, item set $\mathcal{I}$,
  relevance sets $\{\mathcal{O}_u^+ : u \in \mathcal{U}\}$
repeat
  Sample u uniformly from $\mathcal{U}$, i^+ uniformly from $\mathcal{O}_u^+$
  $N = 0$
  violatorFound = False
  repeat
    Sample i^- uniformly from $\mathcal{I} \setminus i^+$
    if $f_\theta(u,i^+) - f_\theta(u,i^-) < 1$ **then**
      violatorFound = True; $v = i^-$
      break
    end if
    $N = N + 1$
  until $N >= \frac{|\mathcal{I}|-1}{\gamma}$
  if violatorFound **then**
    Take gradient step on
    $\mathcal{C}\left(\left\lfloor \frac{|\mathcal{I}|-1}{N} \right\rfloor\right)(f_\theta(u,v) - f_\theta(u,i^+) + 1)$
  end if
until max iterations exceeded

Experiments

Dataset Statistics

Dataset	Users	Items	Interactions	Sparsity
last.FM	10000	10000	97727	0.097%
MovieLens	9888	5000	711084	1.44%
Amazon Games	17437	17915	201154	0.064%

Performance of various measures

	Amazon Games	
	MF-AUC	MF-ADG
ATOP	0.7584 (0.0014)	0.7546 (0.0049)
MAP	0.0104 (0.0003)	**0.0124 (0.0003)**
NDCG	0.1460 (0.0006)	**0.1501 (0.0004)**
rec@10	0.0170 (0.0004)	**0.0211 (0.0007)**
ADG	0.1080 (0.0005)	**0.1110 (0.0003)**

	last.FM	
	MF-AUC	MF-ADG
ATOP	0.7490 (0.0064)	0.7449 (0.0028)
MAP	0.0242 (0.0006)	**0.0281 (0.0006)**
NDCG	0.1701 (0.0007)	**0.1750 (0.0008)**
rec@10	0.0473 (0.0010)	**0.0539 (0.0019)**
ADG	0.1294 (0.0005)	**0.1332 (0.0006)**

	MovieLens	
	MF-AUC	MF-ADG
ATOP	0.8854 (0.0018)	0.7449 (0.0028)
MAP	0.0242 (0.0006)	**0.0281 (0.0006)**
NDCG	0.1701 (0.0007)	**0.1750 (0.0008)**
rec@10	0.0473 (0.0010)	**0.0539 (0.0019)**
ADG	0.1294 (0.0005)	**0.1332 (0.0006)**

Test vs validation performance

	MF-AUC			MF-ADG		
Measure	Test	Valid	Diff%	Test	Valid	Diff%
ATOP	0.8855	0.8849	-0.06	0.8821	0.8817	-0.05
ADG	0.1714	0.1709	-0.29	0.1768	0.1768	-0.00
REC@10	0.0945	0.0945	0.00	0.1025	0.1030	0.49
MAP	0.0775	0.0586	-24.38	0.0858	0.0657	-23.43
NDCG	0.3718	0.2957	-20.47	0.3820	0.3046	-20.42