
Cross-Dataset Recommender System for Overcoming Selection Bias

Anonymous Author(s)

Affiliation

Address

email

Abstract

1 The abstract paragraph should be indented $\frac{1}{2}$ inch (3 picas) on both the left-
2 and right-hand margins. Use 10 point type, with a vertical spacing (leading)
3 of 11 points. The word **Abstract** must be centered, bold, and in point size 12.
4 Two line spaces precede the abstract. The abstract must be limited to one para-
5 graph. The abstract paragraph should be indented $\frac{1}{2}$ inch (3 picas) on both the
6 left- and right-hand margins. Use 10 point type, with a vertical spacing (leading)
7 of 11 points. The word **Abstract** must be centered, bold, and in point size 12. Two
8 line spaces precede the abstract. The abstract must be limited to one paragraph.
9 testf

1 Introduction

11 Selection bias is one of the most prevalent sources of distribution shift for recommender systems.
12 Selection bias happens when there is a pattern in the users' ratings that is unique to the training
13 set, either implicitly or explicitly. For example, in a recommender system for movies, users might
14 only watch and rate movies that they particularly like, while distribution the recommender system
15 is deployed on contains all movies regardless of person al tastes. Tackling the selection bias in a
16 recommender dataset has been a constant challenge in designing recommender algorithms

17 Quantization is a practical way to categorize an input from a continuous distribution to a discrete
18 set of bins. We assume the ground truth of a user's rating is a continuous distribution over items
19 because of the inherent impossibility of numerically expressing the feeling of enjoyment and excite-
20 ment. We treat a recommender dataset as a noised quantization of the ground truth. In other words,
21 users convert their feelings into one of the discrete choices when they rate an item. Recommender
22 datasets are quantized into different degrees. For example, a typical movie rating dataset uses a
23 5-quantization. A tinder like/dislike dataset uses a 2-quantization.

24 We hypothesize that quantization and the effect of selection bias on the recommender system is
25 positively correlated: the more quantized the dataset is (the fewer bins there are), the more severe
26 the selection bias becomes in terms of the recommender system's performance. Therefore, a good-
27 performing recommender system should prioritize training data that is less quantized.

28 In reality, however, less quantized data might only represent a small subset of all the data one is
29 capable of collecting. This is because less quantized data is typically explicit while more quantized
30 data is typically implicit. That is to say, less quantized data would require a user explicitly rating an
31 item, while more quantized data would be more easily gathered by observing the users' behaviors.
32 For example, by observing the movie a user spontaneously chooses to see or if a user watches a
33 movie's trailer to the end, we can easily create a 1-quantized or 2-quantized dataset, correspondingly,
34 without the user paying any effort. In this paper, we propose a novel framework that utilizes these

35 data. Our framework specifically exploits the fact that more quantized data is more susceptible to
 36 selection bias and treats it as an advantage that helps less quantized dataset better overcome the
 37 damage caused by selection bias.

38 The main contribution of this work is two-fold:

- 39 1. We design experiments under a simulated environment and shows that susceptibility to
 40 selection bias is likely not correlated with the way a dataset is quantized.
- 41 2. We propose a cross-dataset learning framework that empowers existing recommender algo-
 42 rithms to make use of the more quantized, less biased data. We carried out experiment and
 43 found our method outperforming baselines by a significant margin.

44 2 Related Work

45 3 Susceptibility to Selection Bias

46 3.1 Simulated Environment for Controlling Bias

47 3.2 Results

48 4 Cross-Dataset Propensity Estimation

$$\arg \min_{V, W, A} \frac{1}{N} \left(\frac{(Y_{u,i} - (V_u^T W_i + A))^2}{P_{u,i}} \right) + c \|A\|^2$$

49 where $A = \{b_u, b_i, \mu\}$ represents the standard bias parameters (offset), $\hat{Y} = V_u^T W_i + A$ is the
 50 predicted rating, and $c \|A\|^2$ is the regularizer.

51 Stochastic gradient descent (SGD) step for one sample rating (i, j) :

$$b_u \leftarrow b_u - \alpha (b_u - \frac{1}{P_{i,j}} (r_{i,j} - \hat{r}_{i,j}))$$

52

$$b_i \leftarrow b_i - \alpha (b_i - \frac{1}{P_{i,j}} (r_{i,j} - \hat{r}_{i,j}))$$

53

$$V_i \leftarrow V_i - \alpha (V_i - \frac{1}{P_{i,j}} W_j (r_{i,j} - \hat{r}_{i,j}))$$

54

$$W_j \leftarrow W_j - \alpha (W_j - \frac{1}{P_{i,j}} V_i (r_{i,j} - \hat{r}_{i,j}))$$

55 where α is the learning rate, which can be tuned independently for each parameter. In our experi-
 56 ments, we fix α across parameters.

57 5 Experiment

58 We designed experiments under simulated environments that verify the results introduced above. We
 59 first verify that more quantized dataset is more susceptible to selection bias and MNAR. We then
 60 runs our learning framework and compares it with multiple baselines.

61 5.1 Dataset and Baselines

62 5.2 Results

63 6 Conclusion

64 Acknowledgements

65 References

66 7 Draft Area

67 (Consider using 2-dimensional Gaussian?)

68 Assume the (flattened) ground truth ratings is a standard Gaussian with differential entropy

$$\begin{aligned}\mathbb{H}(R \sim N(0, 1)) &= \frac{1}{2} \log 2\pi e & (1) \\ &= 2.05 & (2)\end{aligned}$$

69 We can show that quantization (discretization) loses information. Consider a simple binary quanti-
70 zation that splits the ratings around the mean. Denote the quantized r.v. as Q_1 .

$$\mathbb{H}(Q_1) = -2 \cdot \frac{1}{2} \log \frac{1}{2} = 1$$

71 Consider another quantized r.v. Q_2 that divides the distribution four-fold based on the four quantiles.

$$\mathbb{H}(Q_2) = -4 \cdot \frac{1}{4} \log \frac{1}{4} = 2$$

72 We can see that different quantization schemes loses different amount of information.

$$\mathbb{H}(R) > \mathbb{H}(Q_2) > \mathbb{H}(Q_1)$$

73 Next, we can prove that selection bias causes information loss. Assuming the selection bias causes
74 ratings to have unequal probability to be sampled, and higher ratings are more likely to be sampled.
75 The distribution of the observed ratings would then be right-skewed. We prove that, in the binary
76 case, this reduces the information loss. Let p be the probability that the sampled rating is positive.

$$\mathbb{H} = p \log p + (1 - p) \log(1 - p)$$

77

$$\frac{d\mathbb{H}}{dp} = 1 + \log p + \frac{1 - p}{p - 1} - \log(1 - p) = 0$$

78 which is optimized when $p = 0.5$. This result extends to all n -quantizations (can be proved using
79 Lagrange multipliers.) See [http://pillowlab.princeton.edu/teaching/statneuro2018/](http://pillowlab.princeton.edu/teaching/statneuro2018/slides/notes08_infotheory.pdf)
80 [slides/notes08_infotheory.pdf](http://pillowlab.princeton.edu/teaching/statneuro2018/slides/notes08_infotheory.pdf).