Require: Training Data $\mathcal{D} = \{(x_{ij}, y_{ij})\}$. Hyper-parameter γ

1: $g_{ij}^0 \leftarrow 1$
2: **while** $t = 1 : N_{max}$ **do**
3: $y_{ij}^t \leftarrow y_{ij}/g_{ij}^{t-1}$
4: $w_j^t \leftarrow argmin_{w_j} \frac{1}{ln2} ln(1 + e^{-y_{ij}^t f(x_{ij};w_j)})$
5: $g_{ij}^t \leftarrow \frac{\gamma + C_{ij}}{\gamma + \sum_{e \in I_{ij}} f(x_e; w_j^t)}$
6: **return** w_j

Algorithm 1: Learning Parameter w_j

where C_{ij} is the number of interactions (e.g., clicks), and E_{ij} is the expected number of interactions. The interaction count C_{ij} follows a Poisson distribution $Poisson(E_{ij})$. We can use $\sum_{e \in I_{ij}} f(x_e; w_j)$ to estimate the expected number of clicks, where I_{ij} is the set of all impressions of user and type pair $< i, j >$. Under the assumptions, we can prove that $f(x_{ij}; w_j)$ takes the form of logistic regression, and w_j is a vector of weights on none-interaction features, i.e., $f(x_{ij}; w_j) = \frac{1}{1+\exp(-x_{ij}^T w_j)}$. Also note that in equation (2) γ is a hyper-parameter that sets the initial psudo-counts of the personal correction factor g_{ij}.

In the case of estimating α_{ij} for a new or infrequent user, both C_{ij} and E_{ij} are close to 0. Therefore, the affinity scores for such users largely depend on the model f, i.e. $\alpha_{ij} \approx f(x_{ij}; w_j)$. In the case of frequent user, the type affinity scores would be compensated as $C_{ij} > E_{ij} => g_{ij} > 1$ (penalized as $C_{ij} < E_{ij} => g_{ij} < 1$) when the user interact more (less) frequently with the type than expected.

3 LEARNING PARAMETERS

Algorithm 1 outlines the iterative inference process for learning w_j. At first, the correction factors g_{ij} are initialized to 1. Step 3 and 4 learn the best w_j (coefficient) estimate under the fixed g_{ij}. In this learning process logistic loss is used, as f takes the form of logistic regression. We are adapting weighting described in [3] when solve for w_j. Given the new coefficient estimates, our algorithm updates the correction factors (as the expected number of clicks depends on the underlying feature-based model used) by equation (2). The process iterates until desired number of interactions or reached convergence. In this paper, the stopping criteria is defined as $||(w_j^t - w_j^{t-1})||_2 < 1e^{-10}$. In our experiment, the process takes less than 10 iterations to converge.

The labels used in the training data could be the ground truth affinity scores estimated for frequent users. Alternatively, we could use a click (interaction) prediction model as the feature-based affinity model f. In this case, the label is a binary variable denoting whether a user i interacts (e.g., clicks) with the presented content or not, (e.g., predicting whether a user will click on one story). The non-interaction features (x_{ij}) are extracted only from the user and the content themselves. In doing so, we do not need to explicitly define frequent and infrequent users.

4 EXPERIMENT

We have conducted experiment on Snapchat story data. Snapchat is a large social network that also features a Discover page, where stories from numerous publishers and channels are shown to users.

	Poisson-Gamma (ours)	HTR (3-day)	HTR (1-week)
RMSE	0.429	0.705	0.487

Table 1: Comparison on click count prediction task.

	With affinity (ours)	No affinity
NDCG@10	0.364	0.325
Accuracy	0.87	0.86

Table 2: Comparison on click prediction task.

The Discover page displays the cover images and the titles, once clicked the body of the content would be shown. We evaluate the effectiveness of the affinity scores obtained by our model in two tasks.

For click counts prediction, we aim at predicting how many times a user clicking on stories from a certain publisher using the affinity scores. We use $\alpha_{ij}|I_{ij}|$ as the prediction model (treating α_{ij} as probabilities), where $|I_{ij}|$ is the number of impressions). We compare our estimated affinity scores estimated with historical tap ratio. The historical tap ratio (HTR) is defined as the number of clicks divided by the number of impressions within a period of time per publisher. We use 0.25% of the all users from 7/30/2017 to 08/05/2017 for training, and predict the click counts for the same users on 08/06/2017. We use Root-mean-square error (RMSE) as the evaluation metric. For click prediction task, we aim at predicting whether a user will click on a story by using affinity score as an extra feature. We split the same data into 70% training and 30% testing and use xgboost as the prediction model. We use NDCG@10 and accuracy as the metrics.

Table 1 shows the result of the click count prediction task for different methods. We can clearly see that estimation by the proposed Gamma-Poisson model achieves the best result (RMSE of 0.429) compared with historical tap ratio (RMSE of 0.705 and 0.487 where using 3 days' data and one week's data respectively). In addition, using the affinity score can further improve the click prediction task as shown in Table 2.

5 CONCLUSION

In this paper, we propose a Gamma-Poisson model for accurately estimating the user type affinity scores. We conduct experiments on real data from Snapchat to demonstrate the effectiveness of our model. We plan to further evaluate the estimated scores on various other personalization tasks.

REFERENCES

[1] Gediminas Adomavicius and Alexander Tuzhilin. 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering* 17, 6 (2005), 734–749.

[2] Jonathan L Herlocker, Joseph A Konstan, Loren G Terveen, and John T Riedl. 2004. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)* 22, 1 (2004), 5–53.

[3] Gary King and Langche Zeng. 2001. Logistic regression in rare events data. *Political analysis* 9, 2 (2001), 137–163.

[4] Greg Linden, Brent Smith, and Jeremy York. 2003. Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internet computing* 7, 1 (2003), 76–80.