User Type Affinity Estimation Using Gamma-Poisson Model

Fei Wu Pennsylvania State University fxw133@ist.psu.edu Yanen Li Snap Research Yanen.li@snap.com Ning Xu Snap Reserch Ning.xu@snap.com

ABSTRACT

The affinity of a user to a type of items (e.g., stories from the same publisher, and movies of the same genre) is an important signal reflecting the user's interests. Accurately estimating of the user type affinity has various applications in ranking and recommendation systems. For frequent users, simply dividing the number of interactions with content (e.g., clicks) by the number of impressions (e.g., the number of times the content is presented to each user) would be a good estimate. However, such estimates are erroneous for users who have sparse interaction history, (e.g., new users). To alleviate the problem, feature-based approaches aim to learn functions predicting the affinity score using only none-click features, such as user demographics, locations, and interests. Likewise, such approaches do not take full advantage of the interaction history of frequent users.

Motivated by the limitations of the two approaches, we propose a Gamma-Poisson model that aims at utilizing the interaction history of frequent users, as well as leveraging a feature-based model for infrequent users. Our intuition is that we should rely more on the interaction history when estimating affinity for frequent users, and weigh more on feature-based model for infrequent users. We present experimental results on large-scale real-world data in a publisher content clicks prediction task to demonstrate the effectiveness of the proposed method in estimating user type affinity scores.

1 INTRODUCTION

Estimating affinity of a user to a type of content is a critical task, as the user type affinity directly signals the user interests. An accurate estimation of the user's type affinity can benefit various tasks, such as personalized news feed ranking, item (music, movie, and story) recommendation etc.

For users that have sufficiently adequate interaction history (i.e., frequent users), a simple baseline is to divide the number of interactions with the content (e.g., clicks or purchases) by the number impressions (e.g., the number of times that item of the same type is presented to the user). While this simple baseline could be a good estimation for frequent users, such an estimation is often erroneous for users who infrequently interact with the items (e.g., new users). The problem persists for more sophisticated methods, such as item-based collaborative filtering methods [4] as they need interaction histories.

To address these limitations, feature-based approaches (similar to user-based collaborative filtering) [1, 2] learn functions predicting

© 2018 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC BY 4.0 License.

ACM ISBN 978-1-4503-5640-4/18/04.

https://doi.org/10.1145/3184558.3186918

the affinity score using non-click features (e.g., user demographics). As the prediction model does not use the historical interactions as the input, we can predict the affinity scores for users with infrequent or no interaction history (given sufficiently informative non-click features). For example, a model that learns from the purchasing histories of college students to recommend school supplies during the beginning of semesters is very effective for an electronic commerce platform. The model can correspondingly infer that new users who are colleges students may have the same needs, even if the platform has no interaction history for those users. However, one drawback of the feature-based approaches is that they do not take full advantage of the interaction history of frequent users. As a result, users sharing similar non-click features are considered to be similar, regardless of how they actually interact with different types of items.

Motivated by addressing the drawbacks of the above approaches, we propose a Gamma-Poisson model that aims to utilize the interaction history of frequent users, as well as to leverage a feature-based model for infrequent users. Our intuition is that we should rely more on interaction (e.g., clicks) histories when estimating affinity for frequent users, and more on the feature-based model for infrequent users. We conduct experiments on a large scale dataset for a click count prediction on publisher content and demonstrate the effectiveness of our proposed model.

2 GAMMA-POISON MODEL

Estimating user type affinity is to answer how likely a user *i* is going to interact (e.g., click) on an item of type *j*. We denote such tendency as the user type affinity score α_{ij} . We define α_{ij} as:

$$\alpha_{ij} = f(x_{ij}; w_j) \cdot g_{ij} \tag{1}$$

where $x_{ij} \in \mathbb{R}^d$ is a feature vector of dimension d. $f(x_{ij}; w_j)$ is a feature-based prediction model that takes the feature vectors x_{ij} and output affinity scores. And w_j are learned parameters. Example features for a user could be demographic attributes (e.g., age, gender, and occupation), geo location, and interests. Characteristics for the type can also be used as non-interaction features, such as the popularity of the type, an age range of target etc. g_{ij} is a prior correction factor on the interaction (e.g., clicks) histories between user and type pair < i, j >.

Without loss of generality we assume that (1) one user interaction sequence (e.g., clicks of a user to each type) within a time period is a Bernoulli process, and (2) the interaction sequence (e.g., clicks of all users to each type) of all users is also a Bernoulli process. We also assume that the correction factor g_{ij} follows a Gamma distribution, i.e., $g_{ij} \sim Gamma(\alpha = 1, \beta = \frac{1}{\gamma})$. Based on these assumptions, we have the following of our Gamma-Poisson model:

$$g_{ij} = \frac{\gamma + C_{ij}}{\gamma + E_{ij}} = \frac{\gamma + C_{ij}}{\gamma + \sum_{e \in I_{ij}} f(x_e; w_j)},$$
(2)

This paper is published under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

WWW '18 Companion, April 23-27, 2018, Lyon, France

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: $w_j^t \leftarrow argmin_{w_j} \frac{1}{ln^2} ln(1 + e^{-y_{ij}^t f(x_{ij};w_j)})$
5: $g_{ij}^t \leftarrow \frac{\gamma + C_{ij}}{\gamma + \sum_{e \in I_{ij}} f(x_e; w_j^t)}$
6: return <i>w</i> _{<i>j</i>}

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 $\langle i, j \rangle$. 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 q_{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		
Fable 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

- 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.