Models for Content Creation and Active Communities

Abstract

Social media platforms provide tools for their users to post content, and to interact with other users through messaging, sharing, commenting, etc. For these platforms, it is critical to nurture content creation as conversations on content create active communities and provide value (e.g., stay informed, build meaningful relationships, get a job) to users. Understanding user engagement, and in particular content creation/contribution, is thus important for online recommendation systems. In this abstract, we describe how we build models that learn user content creation behavior and how we leverage such models to actively increase value for creators which encourages them to create content more frequently.

1 Introduction

The content ecosystem in a social network (e.g., Facebook, Instagram, LinkedIn, Google+) can be viewed as a two-sided marketplace of content creators and consumers. The newsfeed and notifications interfaces are two channels used by consumers to engage with content from the creators on these networks. There has been plenty of research on modeling approaches to personalize these channels from the consumer perspective [1, 2]. For simplicity, we will refer to these channels as feed and notifications. In this work, we describe a modeling approach to improve the content creator experience with a focus on feedback from their consumer audience.

One key motivation for content creators is to hear from their desired audience. Improving the creator experience, by getting them more feedback from the right consumers, is critical as it will result in more content creators. More creators will result in increased content liquidity in the ecosystem, thus making the consumption experience more valuable for an increasing number of consumers, who will then provide even more feedback to the content creators. If done right, this can be a virtuous cycle.

A creator-focused model can be leveraged to modify the consumer feed ranking to better balance the consumer interests with the creator values. We present two options for using this model to influence item ranking on feed and notifications, along with some preliminary offline validation results.

2 Motivation and Formulation

A content creator on a social network is typically motivated to engage with her intended audience. The various perceptible measures of success for the creator include the number of views, likes, comments and re-shares that his or her content gets. Different networks may have slightly different mechanisms visible to the content creator. We collectively refer to these various signals as "feedback". In general, established content creators would prefer their audience to keep growing in size, while new content creators would love to get some feedback to confirm that their voice is being heard. *If creators find value in the platform, they continue to create more* – this assumption forms the basis for our choice of using creation propensity as a proxy for creator value. A model which better predicts how feedback affects a creator's future creation behavior can be effectively used as a proxy to represent creator interests during feed ranking for consumers.

2.1 Notation and definitions

Let $Y_{i,t}$ be the number of content pieces created at time t and $X_{i,t}$ be the user features for user i at time t. $X_{i,t}$ comprises of:

32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.

- $a_{i,t}$ is the number of feedback items that user i received in last n days (before time t).
- $< a_{i,t}, Si, t >$ are other treatment heterogeneous features (feedback related interaction terms, such as $< feedback, country >$) of user i at time t.
- $V_{i,t}$ is the set of other features of user i at time t.

Then $P(Y_{i,t} > 0 \mid X_{i,t})$ is the probability of user i creating > 0 content pieces at time t given all features. With the above definitions, we use a linear function to model this distribution as follows:

$$p_{i,t} = P(Y_{i,t} > 0 \mid X_{i,t}) = \mu + \gamma^T V_i + \lambda a_{i,t} + \beta^T < a_{i,t}, Si, t > \quad (1)$$

where μ is the intercept, λ, β^T, and γ^T are the coefficients for $a_{i,t}$, <$a_{i,t}, Si, t$> and $V_{i,t}$. The creator utility from incremental feedback, delta P-Create given delta feedback, can be written as $\delta p_{i,t}$ = $\Delta P(Y_{i,t} > 0 \mid X_{i,t}) = P(Y_{i,t} > 0 \mid (a_{i,t} + \Delta a_{i,t}), S_{i,t}, V_{i,t}) - P(Y_{i,t} > 0 \mid a_{i,t}, S_{i,t}, V_{i,t})$. We will use non-linear functions as a next step.

2.2 Modifying feed ranking

The $\delta p_{i,t}$ is a proxy of creator utility, the incremental probability of a specific creator i generating some content if he or she receives higher amount of feedback. If an item k, whose creator is user i, is being shown to a consumer j, then its feed score $FeedScore(k, j)$ can be written as:

$$FeedScore(k, j) = \alpha E[ConsumerUtility(k, j)] + (1 - \alpha)\delta p_{i,t}, \quad (2)$$

where α is a parameter that controls how much we prioritize the consumer utility over creator utility. One can also explore models which make $\delta p_{i,t}$ specific to the consumer j as well, since a creator may care more about some user's feedback more than others.

2.3 Can there be a matching solution?

Instead of picking an ad hoc α, or finding an optimal value (given Equation 2) through line search, there will be greater value to content creators and consumers alike if we formulate this as a matching problem. Due to space constraints, we will only provide a high-level overview of the formulation for this part. We can formulate the optimization problem as one of maximizing consumer side utility such that the creator side utility is delivered to at least some pre-specified level. While the Lagrangian form would look similar to Equation 2, the discontinuity in the representations of $a_{i,t}$ and $S_{i,t}$ necessitates some non-trivial modifications.

3 Offline Evaluations

We use F1 scores, Area under ROC (AUC), and Area under precision-recall (Average precision) curve as primary performance metrics to analyze the accuracy of $p_{i,t}$. Table 1 shows evaluation results when comparing two models with different definitions of t (Model 101 uses t = 1 day, model 200 and 201 use t = 1 week). where Model 101 estimates the probability of a user creating some content in a day, while Model 201 aims to predict users' creation behavior in a longer time window (one week).

Table 1: Offline Model Evaluation Metrics

Model Version	Cohort	F1	AUC	Average Precision
Model 200 (201 W/O key features)	All	0.59	0.64	0.57
Model 201 (Weekly)	All	0.69	0.70	0.75
Model 101 (Daily)	All	0.56	0.71	0.61
Model 101 (Daily)	DAU		0.70	0.58
Model 101 (Daily)	MAU		0.81	0.84

We observe that the weekly prediction task is easier than the daily prediction task. Also, it is much more challenging to predict creation behavior of more active users (e.g., the Daily Active Users (DAUs)) To further evaluate the model performance and identify areas of improvements, we also keep track of the model results segmented by user cohorts (e.g., users with different engagement-levels).

References

[1] Deepak Agarwal, Bee-Chung Chen, Qi He, Zhenhao Hua, Guy Lebanon, Yiming Ma, Pannagadatta Shivaswamy, Hsiao-Ping Tseng, Jaewon Yang, and Liang Zhang. Personalizing Linkedin feed. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1651–1660. ACM, 2015.

[2] Eytan Bakshy, Solomon Messing, and Lada A Adamic. Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239):1130–1132, 2015.