Predicting Session Termination and Retention on X from Fine-Grained Interaction Logs (Student Abstract)

Cai Yang[1], **Kokil Jaidka**[1], **Subhayan Mukerjee**[1], **Yphtach Lelkes**[2]

[1]National University of Singapore
[2]University of Pennsylvania
cai.yang@u.nus.edu; jaidka@nus.edu.sg

Abstract

We study when users end a session on X using high-resolution interaction logs from 215 US participants collected over four weeks. Sessions are defined via data-driven inter-activity gaps, and each session is encoded by fine-grained activity counts and duration (versus a simple "activity ratio" baseline). Fine-grained activity features substantially outperform the activity ratio baseline (C-index ≈ 0.76 vs. 0.62 for future sessions; 0.72 vs. 0.60 for unseen users), indicating that the composition of activity types is a strong predictor of disengagement. At the app level, we analyze retention over early adoption windows and find that the ratio of active activity in the first three days is most predictive of later usage. These results highlight session composition and early on-platform behavior as practical levers for forecasting and mitigating premature drop-off.

Supplementary material: https://osf.io/kp38b

Introduction

Predicting when people disengage from a platform, such as ending a session or abandoning the app altogether, is central to designing healthier online experiences. Addressing this problem requires temporal modeling, engagement analytics, and survival analysis, yet existing work has largely progressed along separate tracks. Temporal modeling studies have focused on segmenting event streams into sessions, but have rarely used these sessions for predictive modeling. Engagement analytics in recommender systems and human-computer interaction have examined how action diversity, pacing, and dwell time relate to satisfaction, retention, and churn, typically using aggregated user-level features. Survival analysis has been applied to return probability across domains, yet it seldom incorporates empirically derived session boundaries grounded in observed behavior.

In this work, we unify these strands through a behavior-centric perspective that treats sessions as first-class objects and asks: **To what extent can the type and intensity of activity within a session predict user engagement, and how does predictive accuracy improve as more data accumulate?** This framing highlights the value of interaction logs for anticipating patterns of online disengagement.

Methodology

Data collection. We analyzed behavioral logs from 215 US participants in a four-week field deployment of a custom-built X client. The application passively recorded all user activities with millisecond timestamps, enabling fine-grained reconstruction of session structures. The study protocol was approved by the Institutional Review Board of our institution.

Session identification. We define a session as the maximal sequence of actions whose inter-activity gaps are below a threshold τ. Following Halfaker et al. (2015), we estimate τ by fitting a Gaussian Mixture Model to the log-transformed inter-activity gap distribution. We compare alternative models and find that a three-component model offers the best fit. We set τ at the intersection where within- and between-components are equally likely, yielding $\tau = 202\,\mathrm{s}$. The dataset contains 15,716 sessions from 215 users, and the mean and median durations are 139 s and 269 s. The median participant produced 36 sessions, where 93% performed at least one active activity, with the remainder browsing passively.

Activity. The logs captured 35 activity types (e.g., timeline browsing, tweet viewing, liking). To ensure robustness, we keep the 21 types observed in at least 10% of participants as separate categories and merge the rest into an "Other" class, yielding 22 event types for sequence modeling. Furthermore, we categorize activity into similar types. For instance, we refer to "Timeline" as all browsing activities on home timeline and latest timeline. We use the grouped activity in the later analysis. We also map activity to passive vs. active groups (e.g., browsing vs. posting). For each session, we compute counts and total time per activity type.

Survival models. Let T_i be the time to termination for unit i (a session or a user) with covariates $\mathbf{x}_i$ (details in Results) and event indicator $\delta_i \in \{0,1\}$. We consider *Cox proportional hazards (Cox PH)* (Cox 1972) and estimate $\boldsymbol{\beta}$ via the partial log-likelihood:

$$h(t \mid \mathbf{x}_i) = h_0(t)\exp(\eta_i), \qquad \eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}.$$

$$\ell(\boldsymbol{\beta}) = \sum_{i:\,\delta_i=1} \Big[\eta_i - \log \sum_{j\in\mathcal{R}(t_i)} \exp(\eta_j)\Big],$$

	Sessions from the Test Set						Sessions from the Validation Set					
	Activity Types			Activity Ratio			Activity Types			Activity Ratio		
	C-index↑	AUC_t↑	Brier↓	C-index↑	AUC_t↑	Brier↓	C-index↑	AUC_t↑	Brier↓	C-index↑	AUC_t↑	Brier↓
Cox PH	$0.759_{\pm\,0.01}$	$0.856_{\pm\,0.01}$	$0.118_{\pm\,0.00}$	$0.621_{\pm\,0.03}$	$0.687_{\pm\,0.04}$	$0.154_{\pm\,0.00}$	$0.721_{\pm\,0.01}$	$0.812_{\pm\,0.01}$	$0.138_{\pm\,0.01}$	$0.607_{\pm\,0.03}$	$0.666_{\pm\,0.04}$	$0.167_{\pm\,0.01}$
Weibull	$0.758_{\pm\,0.01}$	$0.856_{\pm\,0.01}$	$0.119_{\pm\,0.00}$	$0.621_{\pm\,0.03}$	$0.687_{\pm\,0.04}$	$0.154_{\pm\,0.00}$	$0.721_{\pm\,0.01}$	$0.812_{\pm\,0.01}$	$0.140_{\pm\,0.01}$	$0.607_{\pm\,0.03}$	$0.666_{\pm\,0.04}$	$0.167_{\pm\,0.01}$
LN	$0.755_{\pm\,0.01}$	$0.853_{\pm\,0.01}$	$0.121_{\pm\,0.01}$	$0.621_{\pm\,0.03}$	$0.687_{\pm\,0.04}$	$0.154_{\pm\,0.01}$	$0.719_{\pm\,0.01}$	$0.810_{\pm\,0.01}$	$0.142_{\pm\,0.01}$	$0.607_{\pm\,0.03}$	$0.666_{\pm\,0.04}$	$0.170_{\pm\,0.01}$
LL	$0.756_{\pm\,0.01}$	$0.853_{\pm\,0.01}$	$0.122_{\pm\,0.01}$	$0.621_{\pm\,0.03}$	$0.687_{\pm\,0.04}$	$0.155_{\pm\,0.01}$	$0.719_{\pm\,0.01}$	$0.809_{\pm\,0.01}$	$0.143_{\pm\,0.01}$	$0.607_{\pm\,0.03}$	$0.666_{\pm\,0.04}$	$0.169_{\pm\,0.01}$

Table 1: Session-termination performance. C-index: Concordance index. AUC_t: time-dependent AUC. Brier: Brier score. C-index and AUC_t on Activity Ratio remain the same, as there is only one covariate, thus producing the same ranking.

	D=3			D=5			D=7			D=10		
	C-index↑	AUC_t↑	Brier↓	C-index↑	AUC_t↑	Brier↓	C-index↑	AUC_t↑	Brier↓	C-index↑	AUC_t↑	Brier↓
Cox PH	$0.671_{\pm\,0.08}$	$0.651_{\pm\,0.15}$	$0.048_{\pm\,0.01}$	$0.619_{\pm\,0.08}$	$0.619_{\pm\,0.11}$	$0.048_{\pm\,0.01}$	$0.631_{\pm\,0.10}$	$0.621_{\pm\,0.15}$	$0.049_{\pm\,0.01}$	$0.560_{\pm\,0.07}$	$0.503_{\pm\,0.08}$	$0.050_{\pm\,0.01}$
Weibull	$0.668_{\pm\,0.08}$	$0.644_{\pm\,0.14}$	$0.048_{\pm\,0.01}$	$0.619_{\pm\,0.08}$	$0.623_{\pm\,0.10}$	$0.048_{\pm\,0.01}$	$0.670_{\pm\,0.08}$	$0.643_{\pm\,0.15}$	$0.049_{\pm\,0.01}$	$0.605_{\pm\,0.04}$	$0.536_{\pm\,0.11}$	$0.051_{\pm\,0.01}$
LN	$0.668_{\pm\,0.09}$	$0.630_{\pm\,0.15}$	$0.048_{\pm\,0.01}$	$0.612_{\pm\,0.07}$	$0.607_{\pm\,0.10}$	$0.048_{\pm\,0.01}$	$0.636_{\pm\,0.11}$	$0.609_{\pm\,0.16}$	$0.049_{\pm\,0.01}$	$0.597_{\pm\,0.04}$	$0.517_{\pm\,0.10}$	$0.050_{\pm\,0.01}$
LL	$0.668_{\pm\,0.08}$	$0.644_{\pm\,0.14}$	$0.048_{\pm\,0.01}$	$0.615_{\pm\,0.08}$	$0.618_{\pm\,0.10}$	$0.048_{\pm\,0.01}$	$0.662_{\pm\,0.09}$	$0.634_{\pm\,0.16}$	$0.049_{\pm\,0.01}$	$0.595_{\pm\,0.05}$	$0.517_{\pm\,0.09}$	$0.050_{\pm\,0.01}$

Table 2: App-level exit using the Active Ratio over the first $D \in \{3, 5, 7, 10\}$ days.

where $\mathcal{R}(t_i)$ is the risk set at t_i. We also consider *Accelerated failure time (AFT)* (Wei 1992):

$$\log T_i = \mathbf{x}_i^\top \boldsymbol{\gamma} + \sigma\,\varepsilon_i,$$

where we consider Weibull, log-logistic (LL) and log-normal (LN) for error distribution ϵ, with parameters $(\boldsymbol{\gamma}, \sigma)$ estimated by the full log likelihood:

$$\ell(\boldsymbol{\gamma}, \sigma) = \sum_i \Big[\delta_i \log f(t_i \mid \mathbf{x}_i) + (1 - \delta_i) \log S(t_i \mid \mathbf{x}_i)\Big],$$

where f and S are the model-specific density and survival functions.

Results and Discussion

Session termination. Covariates include (i) the activity types and (ii) the activity ratio, defined as the fraction of activities involving active user engagement, such as liking, posting, or replying to others, relative to total session activities. We consider both count- and time-based composition and observe similar trends (same for app exit). For brevity, we report count-based results. We evaluate two setups: *Sessions from the test set*: forecasting time-to-termination for a user's subsequent sessions based on prior sessions (we consider the last S sessions to test where $1 \leq S \leq 5$), and *Session from the validation set*: forecasting unseen users' sessions. Because sessions always end, there is no right-censoring (δ_i=1 for all i). Table 1 displays the models' performance evaluated with 5-fold cross validation using three metrics: concordance index, time-dependent AUC and Brier score. The former two metrics are discrimination-based while the latter measures the overall prediction error. Across both settings, using activity types yields consistently better discrimination and lower errors. Difference between Cox and AFT families is minor. Inspecting model fit reveals that sessions with more passive scrolling (e.g., timeline viewing) tend to end sooner, whereas viewing tweet details and light interactions (e.g., favor, retweet) persist longer.

App exit. We apply survival analysis to app-level retention using the early activity ratio of each user. We consider users with less than 21 days of activity as early-leavers (16 in total), which constitutes 75% of the entire timeframe. To avoid over-parameterization, we use the daily activity ratio as the sole covariate, computed over the first $D \in \{3,5,7,10\}$ days. Table 2 reports performance across D with a 5-fold cross validation. D=3-7 provides higher discrimination and lower error, whereas D=10 degrades across metrics. This indicates that early on-platform behavior carries the strongest signal for subsequent retention.

In summary, our findings illustrate how fine-grained behavioral signals can predict session termination and app-level exit. For future work, we will enrich session features with content composition, such as the topic of tweets, and adopt more advanced survival frameworks. These extensions will support the design and evaluation of interventions to improve user retention.

References

Cox, D. R. 1972. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2): 187–202.

Halfaker, A.; Keyes, O.; Kluver, D.; Thebault-Spieker, J.; Nguyen, T.; Grandprey-Shores, K.; Uduwage, A.; and Warncke-Wang, M. 2015. User Session Identification Based on Strong Regularities in Inter-activity Time. WWW'15, 410–418.

Wei, L.-J. 1992. The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Statistics in medicine*, 11(14-15): 1871–1879.