A Hazard Based Approach to User Return Time Prediction

Komal Kapoor
Jaideep Srivastava

Mingxuan Sun
Tao Ye
Outline

• Motivation
• Techniques
• Data and Findings
• Conclusion
Motivation
User Retention
User Retention

• What works and what doesn’t work?

Redesign and restructuring
User Retention

• What works and what doesn’t work?

Redesign and restructuring

• Identifying at risk users

Selective targeting, promotions, marketing
Retention for the New Web

- Highly dynamic user behavior
  - multiple services in market
  - significant number of drop outs
  - prone to changes
- Engagement is the new metric for success
  - user return rate
  - user time spend
Retention for the New Web

• Highly dynamic user behavior
  • multiple services in market
  • significant number of drop outs
  • prone to changes
• Engagement is the new metric for success
  • user return rate
  • user time spend
Model the Time of Next Return

visitation history

return time

return time

observed length of absence

last day of observation
Techniques
Survival Analysis

• Modeling **time** of occurrence of **event**
  e.g. death, failure, recovery, adoption, **return**, exit, click, etc.

• Handle incomplete (censored) data
  • Users that do not return
    • **Cannot simply discard such users!** (bias)

• Attribute return rate to user features and other events
  • Covariates - feedback, tenure, loyalty
  • Dynamic covariates - time of day, system changes
Cox Proportional Hazard Model

- Measures the rate of user return given the time elapsed
- Shape indicates the dynamics in the return rate
Cox Proportional Hazard Model

Adoption of products

Response to surveys
Cox Proportional Hazard Model

User level features impact the magnitude of the return rate.
Cox Proportional Hazard Model

\[ \lambda(t) = \lambda_0(t) \times \exp(\beta_1 \times X_1(t) + \beta_2 \times X_2(t) + ...) \]
Cox Proportional Hazard Model

\[ \lambda(t) = \lambda_0(t) \times \exp(\beta_1 \times X_1(t) + \beta_2 \times X_2(t) + \ldots) \]

Baseline Hazard Function (non-parametric)
Cox Proportional Hazard Model

\[ \lambda(t) = \lambda_0(t) * \exp(\beta_1 * X_1(t) + \beta_2 * X_2(t) + \ldots) \]

Impact of covariates
Cox Proportional Hazard Model

\[ \lambda(t) = \lambda_0(t) \times \exp(\beta_1 \times X_1(t) + \beta_2 \times X_2(t) + ...) \]
Cox Proportional Hazard Model

\[ \lambda(t) = \lambda_0(t) \times \exp(\beta_1 \times X_1(t) + \beta_2 \times X_2(t) + \ldots) \]

The first and the second terms are independent of each other and are learned separately.
Data and Findings
Music Domain Datasets

- The Last.fm public dataset:
  - 1000 users
  - Training: All user visits during Oct - Dec 2008
  - Testing: All user visits during Jan - Mar 2009

- Large-scale proprietary dataset:
  - 73,465 users
  - Cross Validation: All users visits during May - July 2012

- Multiple observations from the same user are reweighted, each user gets unit weight
Covariates

*Typical visitation patterns of a user*
- Active Weeks
- Density of Visitation
- Visit Number
- Time weighted average return time (TWRT)

*Satisfaction/engagement with the service*
- Duration
- % Distinct Songs
- % Distinct Artists
- % Skips
- Explicit feedback indicators
Baseline Hazard Function

- Baseline hazard has a declining shape
- Indicative of *inertia* (likelihood of return decreases as time spent away increases)
Return Time Prediction

\[ E(\text{Return Time} \mid \text{Model, Covariates}) \]
Return Time Prediction

Weighted Root Mean Squared Error (WRMSE) = \[ \sqrt{\frac{\sum w \times (T_p - T_a)^2}{\sum w}} \]

### WRMSE Return Time Predictions for Last.fm Dataset

<table>
<thead>
<tr>
<th>Method</th>
<th>Training Data (10-fold Cross Validation)</th>
<th>Test Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average</strong></td>
<td>10.55</td>
<td>10.40</td>
</tr>
<tr>
<td>Linear Regression</td>
<td>9.61</td>
<td>9.37</td>
</tr>
<tr>
<td>Decision Tree Regression</td>
<td>9.45</td>
<td>9.15</td>
</tr>
<tr>
<td>Support Vector Machine</td>
<td>10.76</td>
<td>10.33</td>
</tr>
<tr>
<td>Neural Networks</td>
<td>9.58</td>
<td>9.36</td>
</tr>
<tr>
<td><strong>Hazard Based Approach</strong></td>
<td><strong>8.76</strong></td>
<td><strong>8.45</strong></td>
</tr>
</tbody>
</table>

### WRMSE Return Time Predictions for Large-Scale Dataset

<table>
<thead>
<tr>
<th>Method</th>
<th>Training Data (10-fold Cross Validation)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average</strong></td>
<td>18.55</td>
</tr>
<tr>
<td>Linear Regression</td>
<td>18.33</td>
</tr>
<tr>
<td>Decision Tree Regression</td>
<td>18.14</td>
</tr>
<tr>
<td>Support Vector Machine</td>
<td>-</td>
</tr>
<tr>
<td>Neural Networks</td>
<td>18.26</td>
</tr>
<tr>
<td><strong>Hazard Based Approach</strong></td>
<td><strong>16.58</strong></td>
</tr>
</tbody>
</table>
Future Return Time Prediction

\[ E(\text{Return Time} \mid \text{Model, Covariates, Observed Absence}) \]
Future Return Time Prediction

Large-scale WRMSE Future Return Time Prediction

![Graph showing weighted root mean square error over length of absence (LOA) for different prediction methods: Simple Average, Linear Regression, Linear Regression (+LOA feature), Decision Tree, Decision Tree (+LOA features), and Hazard Based Approach. The graph demonstrates how each method performs with varying lengths of absence.]
Classification into User Buckets

- Short Return time
- Long Return time
F-measure for User Classification using LOA (Large-scale Dataset)
Conclusions
Takeaways

• Proposed return time prediction as an approach for improving retention in web services
• Used a Cox proportional hazard model which incorporated dynamic return events and effects of covariates
• Improved performance by using the length of absence (LOA)
• Outperformed state-of-the-art baselines in return time prediction and user classification based on return time
Thanks!