A Hazard Based Approach to User Return Time Prediction

Komal Kapoor
Jaideep Srivastava
UNIVERSITY OF MINNESOTA

Mingxuan Sun
Tao Ye
PANDORA

Outline

- Motivation
- Techniques
- Data and Findings
- Conclusion

Motivation

User Retention

User Retention

- What *works* and what *doesn't work?*

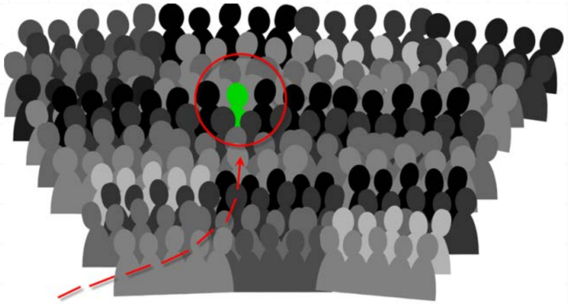
Redesign and restructuring

User Retention

- What *works* and what *doesn't work?*

Redesign and restructuring

- Identifying at risk users



Selective targeting, promotions, marketing

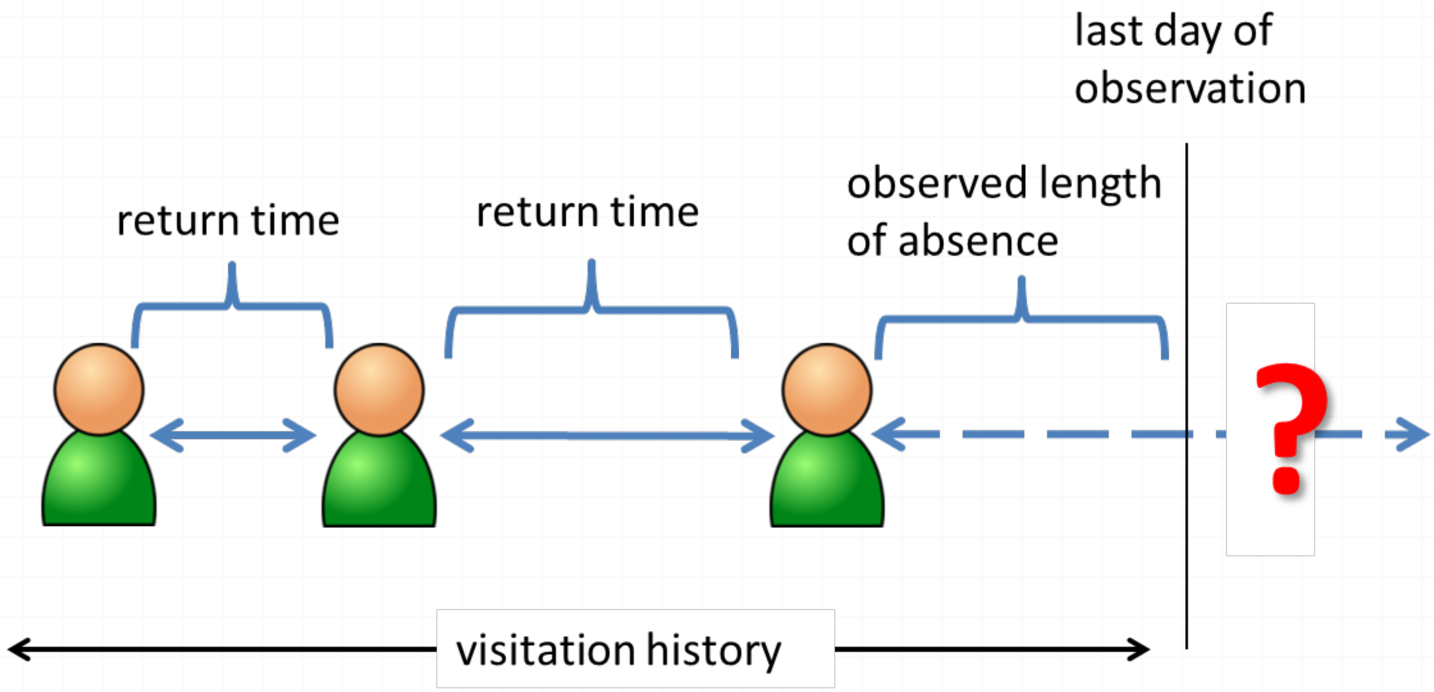
Retention for the New Web

- 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



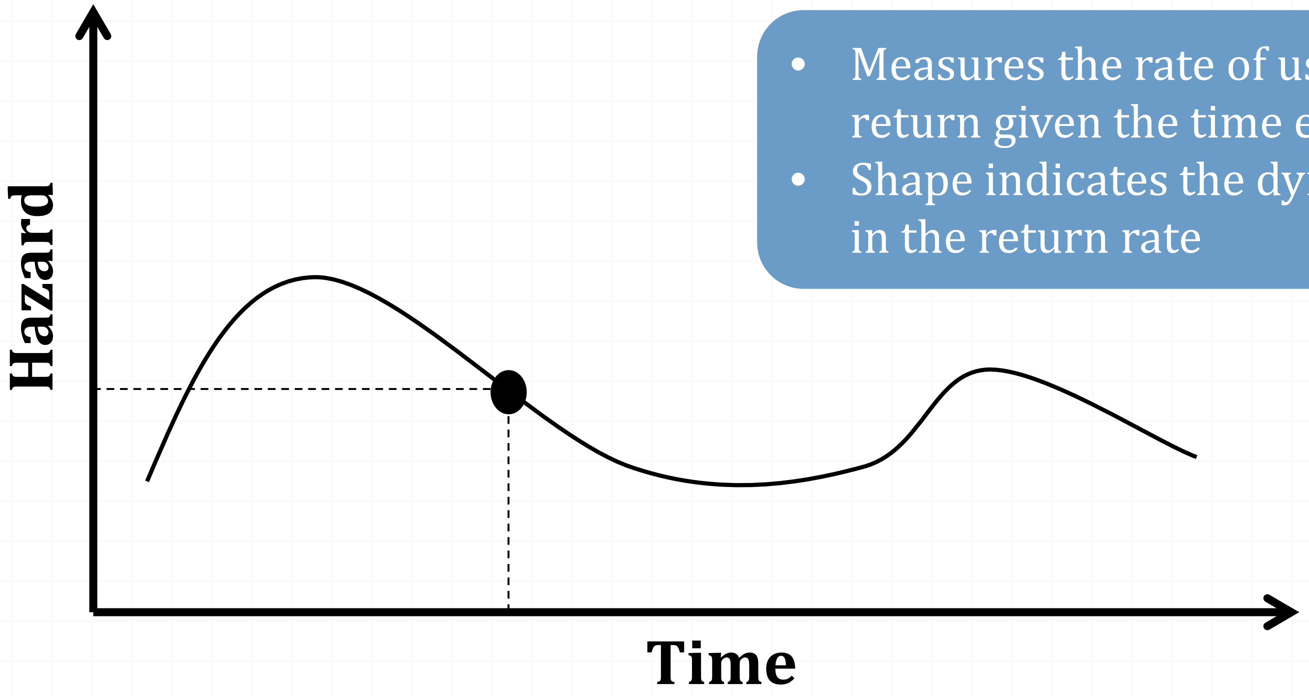
Techniques

Survival Analysis

- Modeling **time** of occurrence of **event**

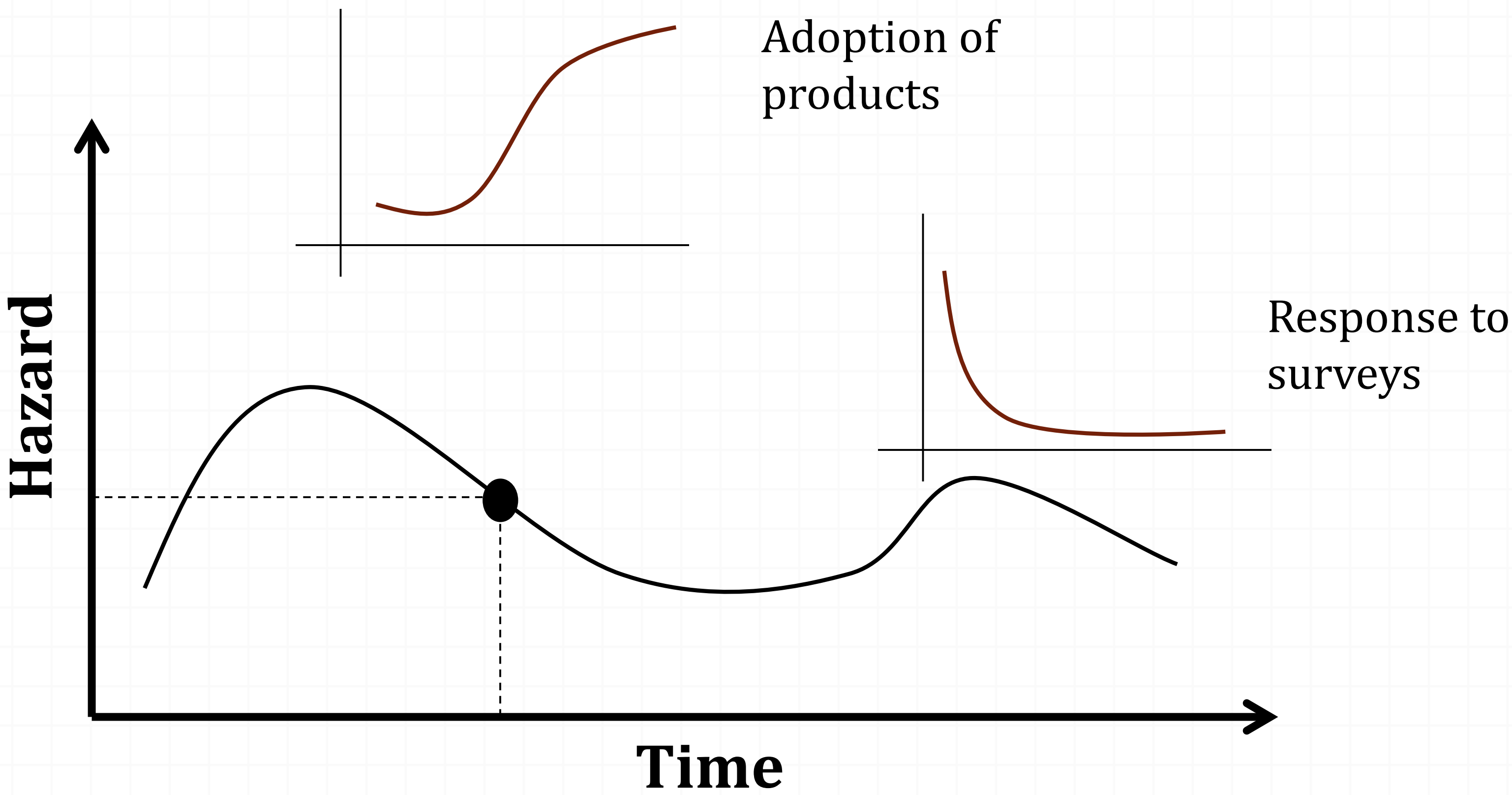
    e.g. death, failure, recovery, adoption, **return,** exit, click, etc.
- Handle incomplete (censored) data
    - Users that do not return
    - Cannot simply discard such users! (bias)
- Attribute return rate to user features and other events
    - Covariates - feedback, tenure, loyalty
    - Dynamic covariates - time of day, system changes

Cox Proportional Hazard Model

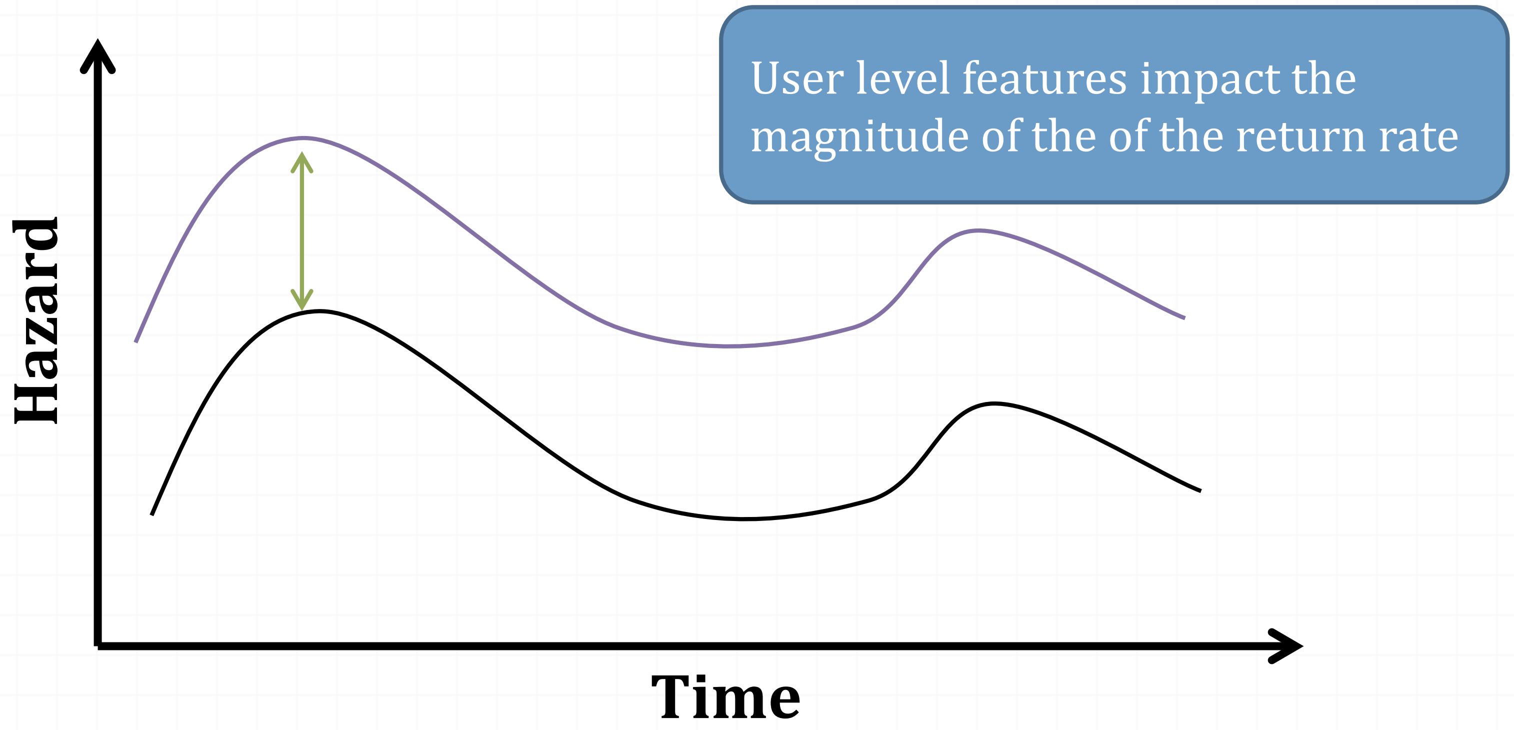


- Measures the rate of user return given the time elapsed
- Shape indicates the dynamics in the return rate

Cox Proportional Hazard Model



Cox Proportional Hazard Model



Cox Proportional Hazard Model

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

Cox Proportional Hazard Model

$$\lambda(t) = \underbrace{\lambda_0(t)}_{\text{Baseline Hazard Function (non-parametric)}} * \exp(\beta_1 * X_1(t) + \beta_2 * X_2(t) + \ldots)$$

Cox Proportional Hazard Model

$$\lambda(t) = \lambda_0(t) * \underbrace{\exp(\beta_1 * X_1(t) + \beta_2 * X_2(t) + \ldots)}_{\text{Impact of covariates}}$$

Cox Proportional Hazard Model

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

Regression Coefficient

Cox Proportional Hazard Model

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

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**

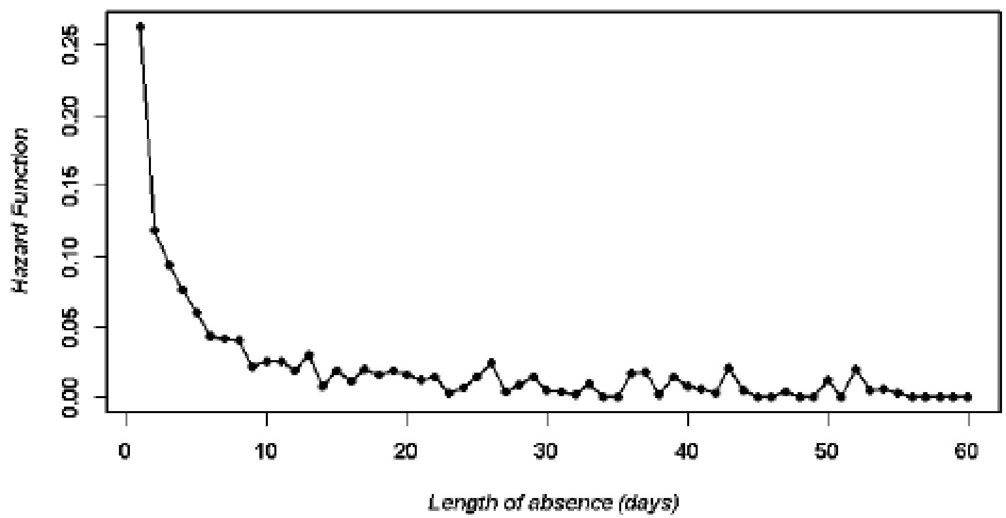
- Active Weeks
- Density of Visitation
- Visit Number
- Time weighted average return time (TWRT)

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(Return Time | Model, Covariates)

Return Time Prediction

Weighted Root Mean Squared Error (WRMSE) = $\sqrt{\frac{\sum w * (T^p - T^a)^2}{\sum w}}$

WRMSE Return Time Predictions for Last.fm Dataset

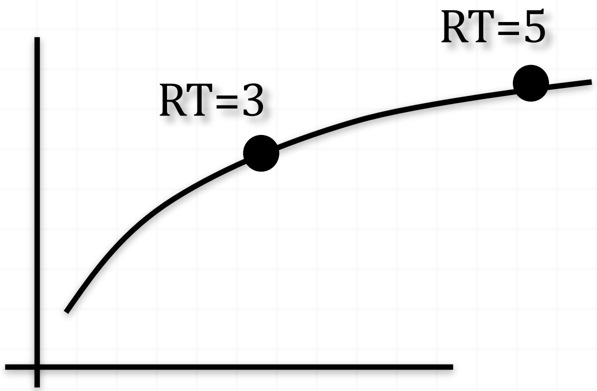
	Training Data (10-fold Cross Validation)	Test Data
Average	10.55	10.40
Linear Regression	9.61	9.37
Decision Tree Regression	9.45	9.15
Support Vector Machine	10.76	10.33
Neural Networks	9.58	9.36
Hazard Based Approach	**8.76**	**8.45**

WRMSE Return Time Predictions for Large-Scale Dataset

	Training Data (10-fold Cross Validation)
Average	18.55
Linear Regression	18.33
Decision Tree Regression	18.14
Support Vector Machine	-
Neural Networks	18.26
Hazard Based Approach	**16.58**

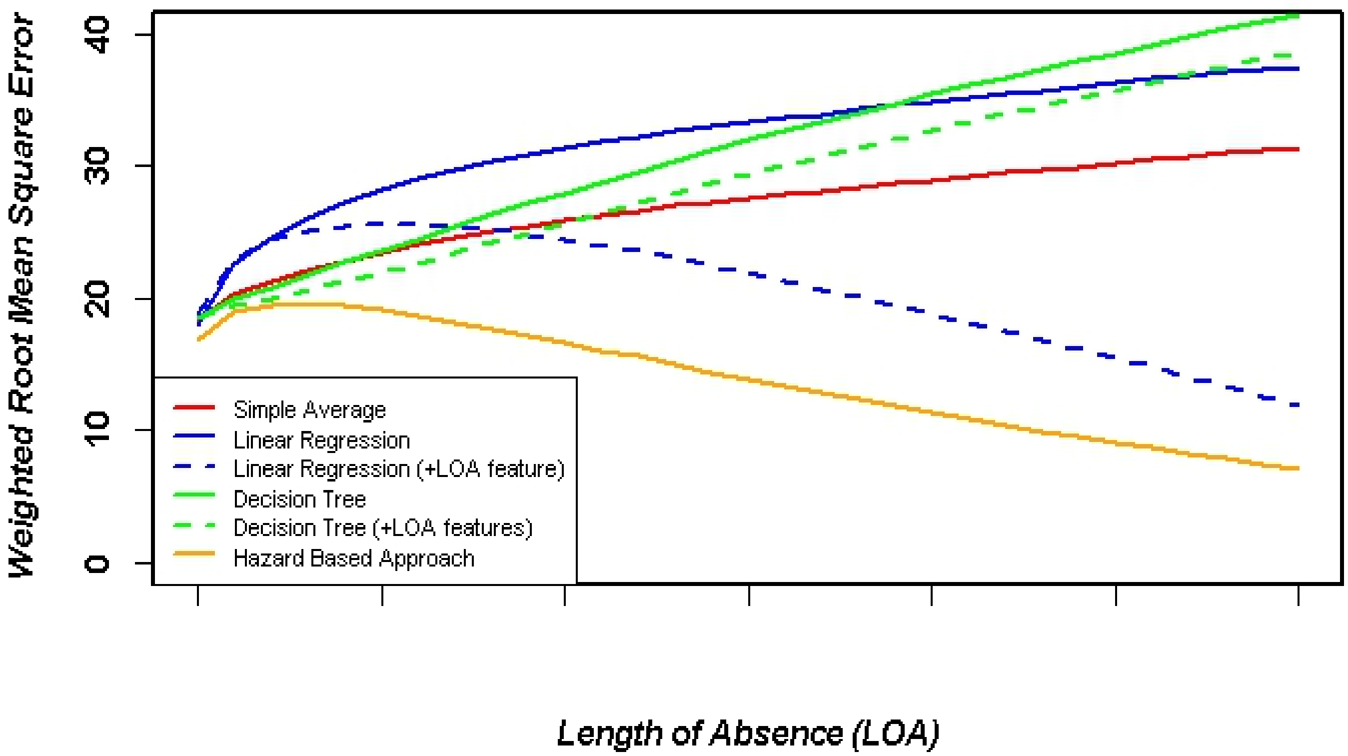
Future Return Time Prediction

E(Return Time | Model, Covariates, **Observed Absence**)



Future Return Time Prediction

Large-scale WRMSE Future Return Time Prediction

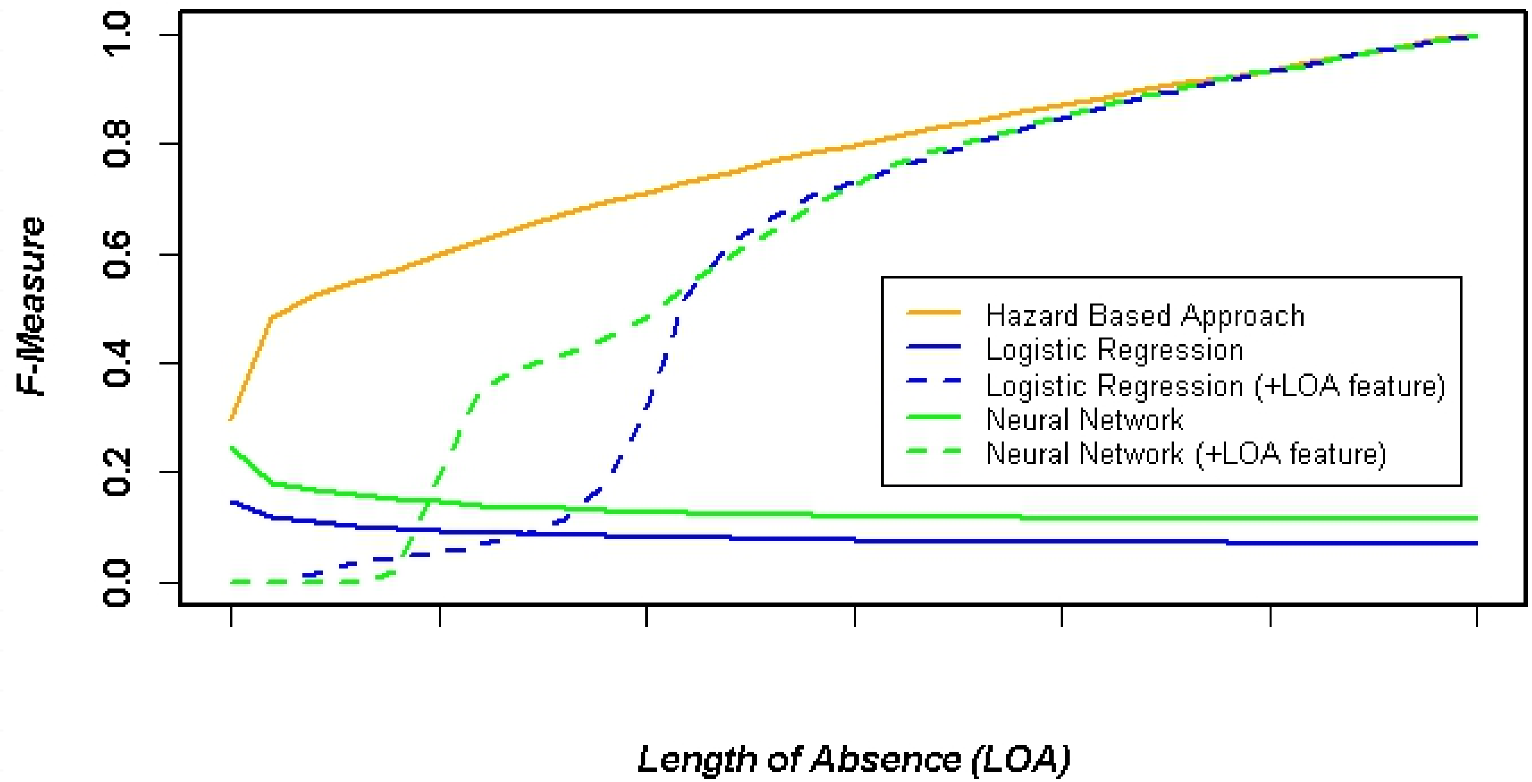


Classification into User Buckets

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!