GeoLifecycle: User Engagement in Geographical Change and Churn Prediction in LBSNs

Presenter: Young D. Kwon
ydkwon@cse.ust.hk

2019. 09

HKUST
Department of Computer Science and Engineering
System and Media Lab
Young D. Kwon, Dimitris Chatzopoulos, Ehsan Ul Haq,
Raymond Chi-Wing Wong, and Pan Hui
Why User Engagement and Churn Prediction?

- Proliferation of Location-Based Social Networks (LBSNs)
  - Heavily rely on User Generated Content (e.g., reviews)
  - Users can stop contributing at any time

Main Goal: Maintainability

Analysis: User Engagement

Application: Churn Prediction

Leave or Stay?

- Yelp Factsheet, August 2018. URL: https://www.yelp.com/factsheet
Challenges

❖ Limitations

• **Unclear** how users engage with LBSNs
  - LBSNs can capture **online** and **offline** experiences of users

• **Effects of various aspects** (e.g., temporal, social, linguistics) are **not fully studied**
  - Novel Offline Feature: Geographic, Venue-specific features
  - New Platform (Not studied yet)

✓ 2013. WWW. No Country for Old Members: User Lifecycle and Linguistic Change in Online Communities. Danescu-Niculescu-Mizil et al.
Challenges

❖ Limitations

- **Less Attention** on churning of highly active *producer-type users* who contribute a majority of reviews

❖ Focus & Scope

- **Focus on** highly active *producer-type users*

- **Limit the scope** of user engagement to reviewing behaviors


✓ 2013. WWW. No Country for Old Members: User Lifecycle and Linguistic Change in Online Communities. Danescu-Niculescu-Mizil et al.

Examined Research Questions

**RQ1:** How do highly active producer-type users engage in the services of LBSNs in terms of *geographical* exploration?

**RQ2:** How do *engagement patterns* of highly active producer-type users manifest themselves in various aspects?

**RQ3:** To what extent can we predict the churning of users with significant contributions within a given period of time?

✓ Yelp dataset: https://www.yelp.com/dataset
RQ1

- **Geographical engagement patterns**: How do highly active producer-type users engage in the services of LBSNs in terms of geographical exploration?

  - $r_g(t)$: The average radius using a user’s trajectory up to $t^{th}$ reviews.
RQ1

- **Human life course:** will users settle down or keep exploring geographically?
  - $d$: Distance to define neighborhoods
RQ2

- How do engagement patterns of highly active producer-type users manifest themselves in various aspects?

**Venue-specific Aspect**

**Linguistic Aspect**

**Social Aspect**
RQ3

- **Churn Prediction Task**: To what extent *can we predict churning* of users with significant contributions within a given period of time?

### Classifiers

1. Logistic Regression (LR) with L2-Regularization
2. Stacked LSTMs

### Models

(F1) Temporal feature (Baseline)
(F2) Geographic feature
(F3) Venue property
(F4) Social feature
(F5) Linguistic feature
(F6) Top2 (based on feature importance)
(F7) Top2+Geo2
(F8) All
(F9:F15) Leave-one-out
RQ3

AUC

# Reviews

- Linguistic
- Geographic
- Venue
- Temporal
- Social
- All-LR
RQ3

The diagram shows the AUC (Area Under the Curve) values for different numbers of reviews. Each bar represents a different feature or combination of features:

- **Linguistic** in red
- **Geographic** in orange
- **Venue** in green
- **Temporal** in blue
- **Social** in brown
- **All-LR** in purple
- **All-LSTMs** in gray

The x-axis represents the number of reviews from 10 to 50, while the y-axis shows the AUC values ranging from 0.50 to 0.90.
Contributions

- Users **constantly wander** around diverse offline places
- The behavioral differences between churners and stayers are significant and are exhibited with their **first 10 reviews**
- LR models based on our findings **significantly improve the performance** over the baseline on the churn prediction task
- We achieve even higher performance in the task by employing a deep learning model
Take-Home Messages

• The average radii and moving distance of users are determined within 5-10 reviews and stable over their lifecycle
  ➢ More personalized services based on a user’s average radius and moving distance

• Users constantly write reviews to diverse locations
  ➢ Recommend to a user different venues located in geographically different neighborhoods that the user have not reviewed yet

• We can accurately predict churning users
  ➢ Gamification techniques such as badges and rewards could be used to increase engagement levels of users
Thank you!

Any questions?

You can find me at:
ydkwon@cse.ust.hk
http://www.youngkwon.org/

Young D. Kwon, Dimitris Chatzopoulos, Ehsan Ul Haq, Raymond Chi-Wing Wong, and Pan Hui
System and Media Lab, Dept. of CSE, HKUST